

10/051-644 09.17.03

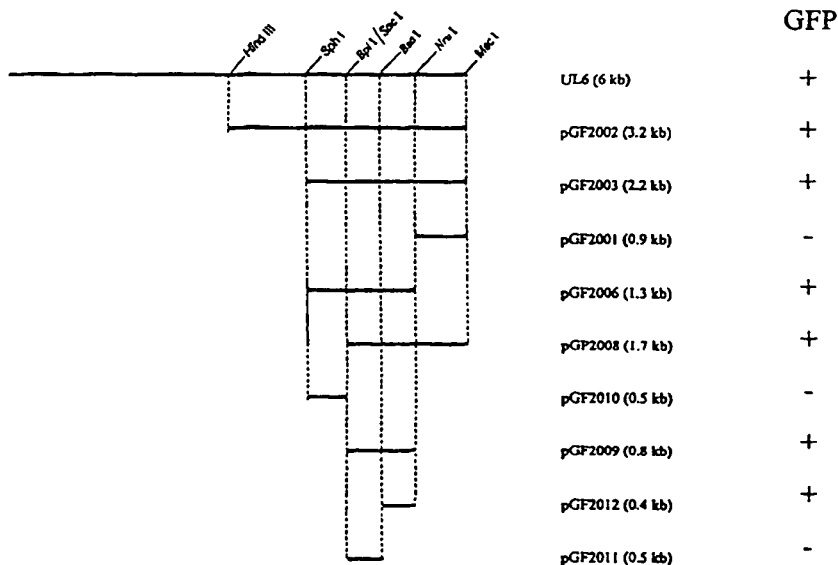


PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : A61K 49/00, C12N 15/12, 15/85, 5/10, C07K 14/435, A01K 67/033, C12Q 1/68		A2	(11) International Publication Number: WO 00/54815
			(43) International Publication Date: 21 September 2000 (21.09.00)
(21) International Application Number: PCT/EP00/02373 (22) International Filing Date: 16 March 2000 (16.03.00) (30) Priority Data: 9906018.8 16 March 1999 (16.03.99) GB (71) Applicant (for all designated States except US): DEVGEN NV [BE/BE]; Technologiepark 9, B-9052 Zwijnaarde (BE). (72) Inventors; and (75) Inventors/Applicants (for US only): ZWAAL, Richard [BE/BE]; (BE). ASAERT, Wouter [BE/BE]; (BE). ROELEN, Ingele [BE/BE]; (BE). BOGAERT, Thierry [BE/BE]; Devgen NV, Technologiepark 9, B-9052 Zwij- naarde (BE). (74) Agents: BAVERSTOCK, Michael, George, Douglas et al.; Boulton Wade Tennant, Verulam Gardens, 70 Gray's Inn Road, London WC1X 8BT (GB).		(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published Without international search report and to be republished upon receipt of that report.	

(54) Title: EXPRESSION OF DNA OR PROTEINS IN *C. ELEGANS*

(57) Abstract

DNA fragments from the promoter region of the *C. elegans* UL6 gene which are capable of functioning as promoters directing gene expression in the excretory cell of *C. elegans* are provided and also expression vectors and transgenic *C. elegans* containing these fragments. Also provided are screening methods performed in *C. elegans* for identifying compounds or mutations which have an effect on the morphology of the excretory canal. Compounds identified using these screening methods may have therapeutic potential in the treatment of a range of diseases for which the *C. elegans* excretory canal serves as a model.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LJ	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

- 1 -

Expression of DNA or proteins in *C. elegans*.

The present invention relates to the expression of DNA, genes, cDNAs, proteins, peptides and parts thereof in the excretory canal of the nematode worm *C.elegans*. In particular, the invention relates to promoter sequences which are capable of directing tissue-specific gene expression in the excretory canal of *C. elegans*, to expression vectors containing the promoter sequences, to transgenic *C. elegans* specifically expressing reporter genes in the excretory canal, to methods of identifying chemical agents that affect the morphology of the excretory canal and to use of these agents in the pharmacological treatment of diseases for which the *C.elegans* excretory canal serves as a model.

The *C.elegans* excretory cell

The excretory system of the nematode *C.elegans* consists of three cells: a single large excretory cell, a duct cell and a pore cell that interfaces with the duct to the main body hypodermis. The excretory cell is the largest mononucleate cell in *C. elegans*. The nucleus and cell body of the excretory cell is situated at the terminal bulb of the pharynx. The cell itself is shaped in an H-form, with the two arms situated along the lateral lines for almost the entire length of the worm, and slightly dorsal. The excretory cell is polarized, having an apical domain facing the lumen of the excretory canal and a basal domain facing outside. The structure and the organization of the *C. elegans* excretory system suggest that it may be used for osmoregulation and can therefore be considered as

- 2 -

a model for the vertebrate nephron.

Various mutant *C. elegans* have been reported which have an aberrant phenotype in the excretory canal. These aberrant phenotypes include cyst
5 formation, short canals and branched canals. Various mutations affecting the excretory canal can be traced back in *C. elegans* II, ed. Riddle, Blumenthal, Meyer and Priess, Cold Spring Harbor Laboratory Press, 1997.

10 Drug discovery in growth cone steering.

Regulation of cell motility, cell shape and the outgrowth of axons or other cell outgrowths are all essential processes in the morphogenesis and function of both unicellular and multicellular organisms.
15 Furthermore, the control of these processes is disturbed in a variety of diseases in which receptors, extra-cellular signals and intra-cellular pathways are over- or under-stimulated. The discovery of new genes, proteins and peptides that are involved in these
20 processes and chemical entities which modulate them would very much help the understanding of these processes. Accordingly, there is a need to develop new methods for the discovery of novel molecules involved in the cell motility, cell shape and cell outgrowth
25 process, and to establish their function. In addition, since malfunction of these biological processes can lead to disease there is also a need to discover chemical entities which modulate these processes which may be useful as pharmaceuticals. Diseases associated
30 with cell motility, cell shape and cell outgrowth include cancerous disease, more particularly tumour formation, metastasis and vascularisation of tumours.

- 3 -

Drug discovery in renal diseases.

In the drug discovery process it is established practice to develop a model of a disease which can be used in the development of assays to screen for compounds with potential pharmaceutical activity. For kidney diseases, and more specifically kidney cyst formation, two different types of disease models currently exist; models based on cell cultures of renal epithelial cells and mouse models. Although these systems have been presented as models for cystic diseases, such as autosomal dominant polycystic kidney disease (ADPKD), they have several disadvantages.

The models based on cell cultures can never be compared with a live multicellular organism. Where aberrant growth indicative of cyst formation has been observed in cultures of different cells, it has proven difficult to develop efficient compound screens from these models. Furthermore, even if chemicals can be discovered that modulate cell growth and hence cyst formation in culture, it remains difficult to prove that these compounds will have analogous effects in the renal systems of multicellular organisms.

The developed mouse models for renal cyst diseases have the disadvantage that they are not suitable for middle to high throughput screening for the discovery of pharmacological compounds. Accordingly, there remains a clear need for an alternative model of renal diseases which more accurately models the renal systems of multicellular organisms but which is practical for use in middle to high throughput screening.

The present invention relates to the use of the *C. elegans* excretory cell in the drug discovery

- 4 -

process. The *C.elegans* excretory canal is an efficient tool to study various developmental biological features; it is formed during the larval stages of the nematode and the canals are observed to grow along the animal in early development. Hence, the development of the excretory canal is an efficient tool to study growth cone steering and defects that might arise during its development and the excretory canal can be used as a model for the development of drug screens in the area of growth cone steering and directional outgrowth.

The *C. elegans* excretory cell and excretory canal can also be considered as a model of the human kidney nephron. The excretory canal has analogous apical-basal polarities as can be found in certain kidney cells and which are relevant for cellular function. Hence, studying the excretory canal may help to develop new tools against kidney diseases. Furthermore, the excretory canal can be used as a model for the development of drug screens in the area of kidney diseases.

In order to exploit the potential of the *C. elegans* excretory cell and excretory canal both as a disease model and in the development of drug screens it would be advantageous to be able to express any gene or cDNA of interest, including reporter genes, specifically in the excretory cell and excretory canal. To achieve this would require the identification of a tissue-specific promoter which is active in the excretory cell.

The present inventors have identified, through the use of biochemical, molecular biology and transgenic techniques, a promoter fragment that

- 5 -

specifically directs transcription in the *C. elegans* excretory cell in a very efficient way. From this promoter fragment several deletions have been generated that still promote transcription, and hence
5 gene expression, in the excretory cell of *C. elegans*. These promoter fragments are useful tools as they can be used to direct specific expression of any DNA fragment of interest in the excretory cell and excretory canal.

10 Accordingly, in a first aspect the invention provides a DNA fragment which is capable of functioning as a promoter directing gene expression in the excretory cell of *C. elegans*, which DNA fragment comprises the sequence of nucleotides set forth in
15 Figure 2 or a fragment thereof in the absence of any other sequence of consecutive nucleotides from the *C. elegans* genome.

The invention further provides a DNA fragment which is capable of functioning as a promoter
20 directing gene expression in the excretory cell of *C. elegans*, which DNA fragment comprises the sequence of nucleotides set forth in Figure 3 or a fragment thereof in the absence of any other sequence of consecutive nucleotides from the *C. elegans* genome.

25 The invention further provides a DNA fragment which is capable of functioning as a promoter directing gene expression in the excretory cell of *C. elegans*, which DNA fragment comprises the sequence of nucleotides set forth in Figure 4 or a fragment
30 thereof in the absence of any other sequence of consecutive nucleotides from the *C. elegans* genome.

The invention further provides a DNA fragment which is capable of functioning as a promoter

- 6 -

directing gene expression in the excretory cell of *C. elegans*, which DNA fragment comprises the sequence of nucleotides set forth in Figure 5 or a fragment thereof in the absence of any other sequence of consecutive nucleotides from the *C. elegans* genome.

The invention further provides a DNA fragment which is capable of functioning as a promoter directing gene expression in the excretory cell of *C. elegans*, which DNA fragment comprises the sequence of nucleotides set forth in Figure 6 or a fragment thereof in the absence of any other sequence of consecutive nucleotides from the *C. elegans* genome.

The invention further provides a DNA fragment which is capable of functioning as a promoter directing gene expression in the excretory cell of *C. elegans*, which DNA fragment comprises the sequence of nucleotides set forth in Figure 7 or a fragment thereof in the absence of any other sequence of consecutive nucleotides from the *C. elegans* genome.

DNA fragments having the nucleotide sequences set forth in Figures 2 to 7 exhibit tissue-specific promoter activity, directing transcription specifically in the excretory cell and excretory canal of *C. elegans*. That is to say when a reporter gene under the control of any one of these DNA fragments is introduced into *C. elegans* a high level of reporter gene expression is observed in the excretory cell and excretory canal with only background expression in other tissues. As will be described below, these tissue-specific promoters are useful tools in the construction of expression vectors which are suitable for directing gene expression specifically in the *C. elegans* excretory cell and excretory canal and in the

- 7 -

construction of transgenic *C. elegans* in which the transgene is expressed specifically in the excretory cell and excretory canal.

5 In the context of the present application, the phrase "in the absence of any further sequence of consecutive nucleotides from the *C. elegans* genome" should be interpreted as meaning in the absence of any other *C. elegans* genomic sequence consecutive with the sequences shown in Figures 2 to 7, respectively. In
10 other words, the DNA fragments of the invention preferably contain the sequences shown in Figures 2 to 7 in the absence of any other consecutive UL6 promoter sequences.

15 In a second aspect the invention provides an expression vector which is suitable for directing tissue-specific expression of a heterologous DNA fragment in the excretory cell of *C. elegans*, the vector comprising a promoter comprising a DNA fragment as set forth in any one of Figures 2 to 7 positioned
20 to direct expression of the heterologous DNA fragment.

The term "heterologous DNA fragment" refers to essentially any DNA fragment which it is desired to express in the excretory cell of *C. elegans*. This DNA fragment can be a gene, a cDNA or a fragment thereof
25 from *C. elegans*, *Drosophila* sp., mouse, human, zebrafish or any other invertebrate or vertebrate origin. Alternatively, the DNA fragment may be of prokaryotic origin, a recombinant DNA or a synthetic DNA fragment. In a preferred embodiment the
30 heterologous DNA is a reporter gene. Suitable reporter genes include those encoding green fluorescent protein (including the many GFP variants and equivalents known in the art), β -galactosidase, β -lactamase, luciferase,

- 8 -

acetoxyhydroxyacid synthase, alkaline phosphatase, β -glucuronidase, chloramphenicol acetyltransferase, horseradish peroxidase, nopaline synthase or octapine synthase. The above are listed by way of example only
5 and it is to be understood that the precise nature of the heterologous DNA fragment is not material to the invention.

In order to achieve expression in eukaryotic host cells (e.g. cells of the nematode worm *C. elegans*) an
10 expression vector must include promoter sequences to position RNA polymerase at the transcription start site and to direct an appropriate frequency of transcription initiation at this site (e.g. to direct tissue-specific expression in the *C. elegans* excretory
15 cell). In accordance with the invention, the promoter region of the expression vector may comprise UL6 promoter sequences which fulfill both functions (i.e. which contain the transcription initiation site for binding of RNA pol and which direct tissue-specific
20 expression) or the promoter region of the vector may comprise a minimal promoter region from an heterologous gene (e.g. the *pes-10* promoter) which functions to position RNA polymerase at the transcription initiation site and possibly to direct a
25 basal level of transcription and UL6 promoter sequences to direct a tissue-specific expression pattern. This will be further understood with reference to the accompanying Examples. The vector might further comprise one or more additional
30 transcriptional regulatory elements (e.g. enhancer elements) in addition to the UL6 promoter sequences.

The expression vector may also include the following elements required for eukaryotic gene

- 9 -

expression: a terminator sequence and downstream polyadenylation signal for transcription termination, translation initiation sequences for ribosome binding, a start codon (usually AUG) and a termination codon for detachment of the ribosome. Such vectors may be obtained commercially or may be assembled from the elements described by methods well known in the art.

Examples of expression vectors according to the invention are plasmids, viral or phage vectors, plasmid vectors being preferred for use in *C. elegans*. Such vectors will normally possess one or more selectable markers, such as a gene for antibiotic resistance. Plasmid vectors, including those designed for expression in *C. elegans*, may also contain a bacterial origin of replication to allow replication in bacterial host cells for cloning purposes. The construction of plasmid vectors suitable for directing expression of a reporter gene in the excretory cell of *C. elegans* are described in detail in the accompanying Examples.

Also provided by the invention are host cells and organisms transformed or transfected with the expression vector.

In a still further aspect the invention provides a transgenic *C. elegans* containing a transgene comprising a promoter which is capable of directing tissue-specific gene expression in the excretory cell of *C. elegans* operatively linked to a protein-encoding DNA fragment.

According to the invention the transgene may comprise any promoter which is capable of directing tissue-specific gene expression in the excretory cell of *C. elegans* operatively linked to any DNA fragment

- 10 -

which it is desired to express in the excretory cell and excretory canal of *C. elegans*. Where a promoter is described herein as being capable of or suitable for directing tissue-specific gene expression in the excretory cell and/or excretory canal this should be taken to mean that the promoter directs a relatively high level of expression in the excretory cell and/or excretory canal and only background expression in other tissues. In a preferred embodiment the promoter comprises a DNA fragment, as described above, comprising a sequence of nucleotides as set forth in any one of Figures 2 to 7 in the absence of any other consecutive sequence of nucleotides from the *C. elegans* genome, i.e. in the absence of any consecutive UL6 promoter sequences. The promoter may, if appropriate, contain sequences from elsewhere in the *C. elegans* genome which are not consecutive with the sequences shown in Figures 2 to 7, for example one or more further cis-acting regulatory elements isolated from a different type of promoter.

In this context the term "transgene" refers to a DNA construct comprising a promoter operatively linked to a protein-encoding DNA fragment. The construct may contain additional DNA sequences in addition to those specified above. The transgene may, for example, form part of a plasmid vector. By the term "operatively linked" it is to be understood that the promoter is positioned to drive transcription of the protein-encoding DNA fragment.

A transgenic *C. elegans* according to this aspect of the invention may be constructed according to any of the standard techniques known to those skilled in the art. A suitable approach involves the construction

- 11 -

of a plasmid-based expression vector in which a protein-encoding DNA of interest is cloned downstream of a promoter capable of directing tissue-specific gene expression in the excretory cell of *C. elegans*.

5 The plasmid vector is then injected into N2 nematodes. In order to facilitate the selection of transgenic nematodes a second plasmid carrying a dominant selectable marker may be co-injected with the experimental plasmid.

10 The plasmid vector is maintained in cells of the transgenic *C. elegans* in the form of an extrachromosomal array. Although plasmid vectors are relatively stable as extrachromosomal arrays they can alternatively be stably integrated into the *C. elegans* genome using standard technology, for example, using
15 gamma ray-induced integration of extrachromosomal arrays (methods in Cell Biology, Vol 48 page 425-480).

The protein-encoding DNA fragment can be any DNA fragment which it is desired to express in the
20 excretory canal of *C. elegans*. This DNA fragment can be a gene, a cDNA or a fragment thereof from *C. elegans*, *Drosophila* sp., mouse, human, zebrafish or any other invertebrate or vertebrate origin. Alternatively, the DNA fragment may be of prokaryotic
25 origin, a recombinant DNA or a synthetic DNA fragment.

In a preferred embodiment the DNA fragment is a promoterless reporter gene encoding a marker protein such as, for example, green fluorescent protein (GFP), β -galactosidase, β -lactamase, luciferase,
30 acetohydroxyacid synthase, alkaline phosphatase, β -glucuronidase, chloramphenicol acetyltransferase, horseradish peroxidase, nopaline synthase or octapine synthase. The expression of a marker such as GFP makes

- 12 -

it possible to visualize the excretory cell *in situ* in the body of the worm without intervening biochemical steps or specialized equipment such as nomarski-microscopy. A GFP-expressing excretory cell can be readily visualized using binocular microscopy following UV irradiation of the worm.

According to the invention, the transgenic *C. elegans* can be of any genetic background, for example, it can be a wild type worm, a selected mutant worm or a transgenic worm. A stably integrated transgene can easily be transferred onto a different genetic background by performing a genetic cross between a first parental *C. elegans* strain containing a stably integrated transgene and a second parental *C. elegans* strain of the desired genetic background. Standard *C. elegans* genetics can be employed for this purpose. The genetic background of the worm generally has no effect on the expression of the transgene in the excretory cell and excretory canal.

In a further embodiment of the invention the transgenic *C. elegans* further comprises a second transgene comprising a promoter which is suitable for directing tissue-specific expression in the excretory cell of *C. elegans* operatively linked to a reporter gene.

This 'double transgenic' *C. elegans* can be constructed by co-injecting *C. elegans* with two plasmid expression vectors; one containing the protein-encoding DNA fragment of interest and the other containing a reporter gene following the procedure described above. Both of the plasmid vectors can be stably integrated into the *C. elegans* genome using standard techniques (methods in Cell Biology,

- 13 -

Vol 48 page 425-480).

In a further aspect the invention provides a method of identifying a mutation in a gene involved in growth cone steering, cell motility, cell shape, tumour formation, metastasis, vascularisation of tumours, the development of the excretory canal, cytoskeletal organisation, surface to cytoskeleton signalling, renal development or kidney disease, which method comprises contacting a transgenic *C. elegans* which expresses a reporter gene in the excretory canal with a mutagen and screening for phenotypic changes in the excretory canal.

Suitable mutagens for use in the method of the invention include EMS, X-rays or the UV-TMP method, all of which are known to those skilled in the art.

Following contact with mutagen the transgenic *C. elegans* are maintained in culture for at least two subsequent generations during which time observations of the morphology of the excretory canals of the progeny are made in order to identify any mutants with an abnormal excretory canal phenotype. Suitable culture conditions are described in the examples given herein.

The expression of a reporter gene such as GFP in the excretory canal allows mutations that affect the development, shape, growth direction and outgrowth of the excretory canal to be observed and selected. The affected gene is then isolated and characterized using standard genetic and molecular biology techniques.

In a preferred embodiment of the method of the invention the transgenic *C. elegans* which express a reporter gene in the excretory canal contains a

- 14 -

transgene comprising a promoter which is suitable for directing tissue-specific gene expression in the excretory canal of *C. elegans* operatively linked to a reporter gene. The promoter preferably comprises a DNA
5 fragment having the sequence of nucleotides set forth in any one of Figures 2 to 7.

The method of the invention may also be adapted for use in the identifying further components of a biochemical pathway involved in growth cone steering,
10 cell motility, cell shape, tumour formation, metastasis, vascularisation of tumours, the development of the excretory canal, cytoskeletal organisation, surface to cytoskeleton signalling, renal development or kidney disease. In this case the
15 mutagenesis method as described above is performed on a transgenic *C. elegans* expressing a reporter gene in the excretory cell whose genetic background is a selected mutant strain. The selected mutant strain is a strain carrying a defined mutation in a gene
20 involved in growth cone steering, cell motility, cell shape, tumour formation, metastasis, vascularisation of tumours, the development of the excretory canal, cytoskeletal organisation, surface to cytoskeleton signalling, renal development or kidney disease.

25 The selected mutant strain may be the result of a previous round of random mutagenesis performed on a wild-type *C. elegans* strain or it may be a known mutant strain, for example a knock-out mutant or an over-expressing strain taken from a *C. elegans*
30 collection (e.g. the *C. elegans* mutant collection at the *C. elegans* Genetic Center, University of Minnesota, St Paul, Minnesota, USA). Methods for creating mutant worms with mutations in selected *C.*

- 15 -

C. elegans genes are known in the art, for example see J. Sutton and J. Hodgkin in 'The Nematode *Caenorhabditis elegans*' Ed. by William B. Wood and the Community of *C. elegans* Researchers CSHL, 1988 594-595; Zwaal et al. 'Target-Selected Gene Inactivation in *Caenorhabditis elegans* by using a Frozen Transposon Insertion Mutant Bank' 1993, Proc. Natl. Acad. Sci. USA 90 pp7431-7435; Fire et al. 'Potent and Specific Genetic Interference by Double-Stranded RNA in *Caenorhabditis elegans*' 1998, Nature 391 860-811.

Further treatment of a selected mutant strain with mutagen results in the production of a double mutant but it is the phenotype of the later mutation (i.e. that resulting from contact with mutagen) which is scored by screening the subsequent generation for further morphological changes in the excretory canal. If the phenotype of the selected mutant is enhanced in the progeny after mutagenesis this indicates that the second mutation has occurred in a gene which acts on the same or a parallel biochemical pathway to the gene affected by the defined mutation. Alternatively, if the phenotype of the selected mutant is suppressed in the double-mutant progeny this indicates that the second mutation event has occurred in an important gene in the biochemical pathway.

Transgenic *C. elegans* for use in this method, or for use in any of the subsequently described methods which require the use of transgenic *C. elegans* expressing protein encoding DNAs and/or reporter genes in the excretory cell and excretory canal, may be constructed according to standard techniques known in the art such as the methodology described by Craig Mello and Andrew Fire, Methods in Cell Biology, Vol 48

- 16 -

Ed. H. F. Epstein and D. C. Shakes, Academic Press,
pages 452-480.

The present invention further provides a method
5 of determining whether a compound is an inhibitor or
an enhancer of the regulation of growth cone steering,
cell motility, cell shape, tumour formation,
metastasis, vascularisation of tumours, renal
development, pathways involved in kidney disease,
10 development of the excretory canal, cytoskeletal
organisation or surface to cytoskeleton signalling,
which method comprises contacting a sample of the
compound with a transgenic *C. elegans* expressing a
reporter gene in the excretory canal and screening for
15 phenotypic changes in the excretory canal.

Contact with a compound which is an inhibitor or
an enhancer of the regulation of growth cone steering,
cell motility, cell shape, tumour formation,
metastasis, vascularisation of tumours, renal
20 development, pathways involved in kidney disease,
development of the excretory canal, cytoskeletal
organisation or surface to cytoskeleton signalling
results in changes the morphology of the excretory
canal. The expression of a reporter gene in the
25 excretory canal allows these changes in morphology to
be easily visualized. Commonly observed abnormal
excretory canal morphologies include: nematodes having
shorter or longer canals, nematodes having curved or
extra branched canals, nematodes having ventral or
30 dorsal canals, nematodes having more or less than two
canals, nematodes having wrongly branched canals,
nematodes having vacuoles or cysts, nematodes with
unusual features in the excretory canal.

- 17 -

Preferred reporter genes include those encoding green fluorescent protein, β -galactosidase, β -lactamase, luciferase, acetohydroxyacid synthase, alkaline phosphatase, β -glucuronidase, chloramphenicol acetyltransferase, horseradish peroxidase, nopaline synthase or octapine synthase. In a preferred embodiment the transgenic *C. elegans* which express a reporter gene in the excretory canal contains a transgene comprising a promoter which is suitable for directing tissue-specific gene expression in the excretory canal of *C. elegans* operatively linked to a reporter gene. The promoter preferably comprises a DNA fragment having the sequence of nucleotides set forth in any one of Figures 2 to 7. The genetic background of the transgenic worm can be wild-type, alternatively the worm can be a mutated worm or a worm expressing a second transgene.

For the avoidance of doubt, it is hereby stated that although the above-described method, and similar screening methods described hereinbelow, are based on bringing *C. elegans* worms into contact with compounds which may potentially have useful pharmacological activity there is no intention to provide any therapeutic benefit to the *C. elegans* during the screen. The worms are used merely as a tool to find out something about the properties of a compound in a biological system. In principle, this is similar to carrying out compound screening *in vitro* using mammalian cells except that the biological system is a microscopic multicellular organism rather than a single cell.

It will be appreciated that a wide variety of candidate compounds may be tested using the screening

- 18 -

methods described herein. The compound may be of any chemical formula and may be one of known biological or pharmacological activity, a known compound without such activity or a novel molecule such as might be present in a combinatorial library of compounds.

The invention further provides a compound which is identifiable using the above method as an inhibitor or an enhancer of the regulation of growth cone steering, cell motility, cell shape, tumour formation, metastasis, vascularisation of tumours, renal development, pathways involved in kidney diseases, development of the excretory canal, cytoskeletal organisation or surface to cytoskeleton signalling.

The invention also provides use of this compound as a medicament, or in the manufacture of a medicament, for promoting neuronal regeneration, re-vascularisation or wound healing or for the treatment of chronic neuro-degenerative diseases, tumour formation, metastasis, tumour vascularisation, kidney diseases, polycystic kidney diseases (specifically ADPKD), cell migration diseases or immunological diseases. Also provided by the invention is a pharmaceutical composition comprising the compound plus a pharmaceutically acceptable carrier, diluent or excipient.

The present invention further provides a method of determining the function of the protein encoded by a DNA fragment, which method comprises the steps of;

(a) providing a transgenic *C. elegans* containing a transgene comprising a promoter which is suitable for directing tissue-specific gene expression in the *C. elegans* excretory cell operatively linked to the

- 19 -

said DNA fragment; and

(b) observing any phenotypic changes in the excretory canal of the transgenic *C. elegans*.

5 According to the method of the invention the function of a given protein or peptide may be studied by expressing the protein in the excretory canal of *C. elegans*. Analysis of any resultant phenotypic changes in the excretory canal may result in a better
10 understanding of the function of the protein.

 The transgene comprises a promoter which directs transcription specifically in the *C. elegans* excretory cell and excretory canal. The DNA fragment encoding the protein of interest, which DNA fragment may be a
15 genomic DNA, a cDNA or a fragment thereof, is placed under the control of this promoter and thereby expressed specifically in the excretory cell and excretory canal. In a preferred embodiment the promoter comprises a DNA fragment having the sequence
20 of nucleotides set forth in any one of Figures 2 to 7.

 In order to help visualise any phenotypic changes in the excretory canal the transgenic *C. elegans* may further comprise a second transgene comprising a promoter suitable for directing tissue-specific gene
25 expression in the *C. elegans* excretory cell operatively linked to a reporter gene, preferably green fluorescent protein, β -galactosidase, β -lactamase, luciferase, acetohydroxyacid synthase, alkaline phosphatase, β -glucuronidase, chloramphenicol
30 acetyltransferase, horseradish peroxidase, nopaline synthase or octapine synthase. As described above, expression of a reporter gene such as GFP makes it possible to visualise the excretory canal without the

- 20 -

need for specialized equipment.

If expression of a given protein in the excretory cell results in an abnormal excretory canal phenotype then in a further embodiment the method of the invention may be combined with a mutagenesis step in order to identify further components of the biochemical pathway on which the protein acts. In this embodiment a transgenic *C. elegans* expressing the protein in the excretory canal is contacted with a mutagen, for example, EMS, UV-TMP or X-rays, all of which are well known in the art, and then maintained in culture for at least two subsequent generations during which time the morphology of the excretory canal is observed. If contact with mutagen results in either enhancement or suppression of the abnormal excretory canal phenotype then this indicates that a mutation has occurred in a gene encoding a component of the same biochemical pathway to that on which the original protein acts or of a parallel biochemical pathway. The mutated gene can then be isolated and characterised using standard molecular biology and biochemical techniques.

The invention further provides a method of determining whether a compound is a modulator of growth cone steering, cell shape, cell motility, tumour formation, metastasis, vascularisation of tumours, renal development, a pathway involved in kidney disease, development of the excretory canal, cytoskeletal organisation or surface to cytoskeleton signalling, which method comprises the steps of;

a) contacting a sample of the compound with a transgenic *C. elegans* expressing a DNA fragment

- 21 -

encoding a protein involved in the regulation of growth cone steering, cell shape, cell motility, renal development or a pathway involved in kidney disease, which transgenic *C. elegans* contains a transgene comprising a promoter which is suitable for directing tissue-specific gene expression in the *C. elegans* excretory cell operatively linked to the said DNA fragment; and

(b) screening for phenotypic changes in the excretory canal.

The transgenic *C. elegans* may be a wild type strain or a selected mutant strain. In one embodiment the transgenic worm has an abnormal excretory canal phenotype. Alternatively, the DNA fragment expressed in the excretory cell of the transgenic worm rescues an abnormal excretory canal phenotype which is present in the genetic background of the transgenic *C. elegans*.

Visible phenotypic changes in excretory canal morphology may include: nematodes having shorter or longer canals, nematodes having curved or extra branched canals, nematodes having ventral or dorsal canals, nematodes having more or less than two canals, nematodes having wrongly branched canals, nematodes having vacuoles or cysts, nematodes with unusual features in the excretory canal. To assist in visualising any phenotypic changes the transgenic worm may further express a reporter gene, preferably encoding green fluorescent protein, in the excretory canal.

The present invention further provides a compound which is identifiable as a modulator of growth cone

- 22 -

steering, cell shape, cell motility, renal development, a pathway involved in kidney disease, development of the excretory canal, cytoskeletal organisation or surface to cytoskeleton signalling according to the above method. The invention also provides use of this compound as a medicament, or in the manufacture of a medicament, for promoting neuronal regeneration, re-vascularisation or wound healing, or for the treatment of chronic neuro-degenerative diseases, metastasis, kidney diseases, kidney cyst formation, polycystic kidney diseases (specifically ADPKD), cell migration diseases or immunological diseases. The compound may be provided as a pharmaceutical composition comprising the compound and a pharmaceutically acceptable carrier, diluent or excipient.

In a still further aspect the invention provides a method of identifying further components of a biochemical pathway on which a compound identifiable as a modulator of growth cone steering, cell shape, tumour formation, metastasis, vascularisation of tumours, cell motility, renal development, a pathway involved in kidney disease, development of the excretory canal, cytoskeletal organisation or surface to cytoskeleton signalling may act, which method comprises the steps of:

- (a) providing a transgenic *C. elegans* which expresses a reporter gene in the excretory canal;
- (b) contacting the transgenic *C. elegans* with a mutagen;
- (c) contacting the mutated *C. elegans* with a compound which is identifiable as a modulator of

- 23 -

growth cone steering, cell shape, tumour formation, metastasis, vascularisation of tumours, cell motility, renal development, a pathway involved in kidney disease, development of the excretory canal,

5 cytoskeletal organisation or surface to cytoskeleton signalling; and

(d) screening for phenotypic changes in the excretory canal.

Following contact with mutagen and the compound
10 the transgenic *C. elegans* are maintained in culture for at least two subsequent generations during which time observations of the morphology of the excretory canals of the progeny are made in order to identify any changes in excretory canal phenotype. As with the
15 other screening methods hereinbefore described the expression of a reporter gene, such as GFP, makes it easy to visualise any phenotypic changes in the excretory canal.

The compound used in this method is one which is
20 identifiable as a modulator of growth cone steering, cell shape, tumour formation, metastasis, vascularisation of tumours, cell motility, renal development, a pathway involved in kidney disease, development of the excretory canal, cytoskeletal
25 organisation or surface to cytoskeleton signalling using one of the methods hereinbefore described for that purpose. As described above, treatment of *C. elegans* with such a compound produces morphological abnormalities in the excretory canal. Suitable
30 mutagens for use in the method of the invention include EMS, X-rays or the UV-TMP method, all of which are known to those skilled in the art.

The use of mutagenesis in the method of the

- 24 -

invention facilitates the identification of further components of the biochemical pathway on which the compound acts. If the mutagenesis step results in the production of mutant progeny in which the abnormal excretory canal phenotype is enhanced (as compared with the phenotype observed following treatment of non-mutated *C. elegans* with the same compound) then the mutation has occurred in a gene in the same biochemical pathway as that on which the compound acts or a parallel pathway. Alternatively, treatment of *C. elegans* according to the method results in mutant progeny in which the abnormal excretory canal phenotype is suppressed, indicating that a mutation has occurred in a gene having an important function on the biochemical pathway on which the compound acts. In either case the gene affected by the mutation can be isolated and characterised using standard molecular biology and biochemical techniques.

In a preferred embodiment of the method of the invention the transgenic *C. elegans* which express a reporter gene in the excretory canal contains a transgene comprising a promoter which is suitable for directing tissue-specific gene expression in the excretory canal of *C. elegans* operatively linked to a reporter gene. The promoter preferably comprises a DNA fragment having the sequence of nucleotides set forth in any one of Figures 2 to 7. The genetic background of the transgenic *C. elegans* may be wild type or it may be a selected mutant strain.

30

The invention will be further understood with reference to the following Examples, together with the accompanying Figures in which:

- 25 -

Figure 1 shows the nucleic acid sequence of the insert of pUL6#64A1.

5 Figure 2 shows the nucleic acid sequence of the insert of pGF2002.

Figure 3 shows the nucleic acid sequence of the insert of pGF2003.

10 Figure 4 shows the nucleic acid sequence of the insert of pGF2006.

Figure 5 shows the nucleic acid sequence of the insert of pGF2008.

15 Figure 6 shows the nucleic acid sequence of the insert of pGF2009.

20 Figure 7 shows the nucleic acid sequence of the insert of pGF2012.

Figure 8 shows a restriction map of the UL6 fragment.

25 Figure 9 is an overview of the series of plasmids containing deletion fragments of UL6.

Figure 10 is a plasmid map of pGF2006.

30 Figure 11 is a plasmid map of pGF2009.

Figure 12 is a plasmid map of pGF2013.

Figure 13 is a plasmid map of pGF2014.

- 26 -

General Experimental methods.

Sequence information and expression patterns were obtained from the *C. elegans* genome project, the Sanger Centre, and Washington University School of medicine, <http://www.sanger.ac.uk/projects/C-elegans>.

All Molecular biology work was performed using standard techniques known in the art, as described by Sambrook et al. (1989), *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press; F. M. Ausubel et al. (eds.) or *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc. (1994), or using minor modifications of the methods described therein.

All manipulations of *C.elegans* worms were performed using techniques described in *Methods in Cell Biology*, vol 84; *Caenorhabditis elegans: modern biological analysis of an organism*, ed. Epstein and Shakes, academic press, 1995, or using minor modifications of the methods described therein.

Transgenic *C. elegans* strains were constructed by injection of plasmid DNA into N2 worms using standard techniques known in the art (see *Methods in Cell Biology*, vol 84 as mentioned above). In order to facilitate the selection of transgenic strains the plasmid pRF4 (Mello, C. C. et al. EMBO J. 10, 3959-3970 (1991)) which carries the rol-6 gene was co-injected with the experimental plasmids as a marker. *C. elegans* expressing rol-6 exhibit the roller phenotype. Any other *C. elegans* dominant selectable phenotypic marker could be used in place of rol-6 with equivalent effect. When generating transgenic *C. elegans* strains with plasmids that encode for and

- 27 -

express GFP or GFP fusion proteins, no co-injection with a dominant selective marker is needed, as the transgenic lines can be selected by simple isolation of the progeny that expresses GFP.

5

Example 1

GFP expression in the excretory canal using a minimal promoter fragment.

pUL#64A1 was isolated as the result of a promoter trapping experiment described by Young J.M. and Hope I.A. Molecular markers of differentiation in *Caenorhabditis elegans* obtained by promoter trapping (1993) *Dev. Dyn.*, **196**:124-132. In this study partial Sau3AI restriction enzyme fragments of *C.elegans* genomic DNA were cloned in the *Bam*HI restriction site of the vector pPD22.11 (described by Fire A, Harrison S.W., and Dixon D. A modular set of LacZ fusion vectors for studying gene expression in *Caenorhabditis elegans* (1990) *Gene* **93**:189-198.) creating LacZ translational fusions. Introduction of pUL#64A1 into *C.elegans* resulted in the expression of β -galactosidase in the excretory cell and excretory canal and the lateral nuclei of the hypodermis adjacent to the anterior and posterior of the excretory cell. The region of the genomic DNA insert immediately adjacent to the lacZ gene was sequenced enabling the origin of the insert of pUL#64A1 to be localized on the *C. elegans* physical genome map.

In order to determine the length of the cloned insert the vector pUL#64A1 was digested with several restriction enzymes, including *Xho*I, *Sal*I, *Sph*I and *Hind*III. Restriction fragment length analysis showed

- 28 -

that the cloned fragment in pUL#64A1 was approximately 6 kb in length. Analysis of the genomic DNA of *C.elegans* (available from the *C.elegans* genome project, the Sanger Centre, and Washington University School of medicine, <http://www.sanger.ac.uk/projects/C-elegans>) revealed the presence of two adjacent *Sau3AI* in the region of the pUL#64A1 insert. This confirmed that the inserted DNA fragment had a length of 6065 bp or 6023 bp and could be located between positions 33561 and 39620 on cosmid C17H12. The inserted DNA fragment was designated UL6.

Analysis of the nucleotide sequence of the UL6 insert revealed the presence of two putative genes orientated in opposite directions. The putative promoter located upstream of the gene orientated in the same direction as the *lacZ* reporter gene was considered to be the most relevant for further analysis.

In order to test for promoter activity several deletion fragments of UL6 spanning this region were cloned in the GFP-fusion vector pPD95.79 (constructed by Andrew Fire, Carnegie Institute of Washington and freely available at ftp://stein.cshl.org/pub/elegans_vector/). The promoter activity of each of the fragments was then tested by injecting the plasmids into *C. elegans* and analysing the levels of GFP expression in different tissues.

A first construct, designated pGF2002, contained the 3.2 kb *HindIII*-*MscI* fragment of UL6. Following injection of pGF2002 into *C. elegans* worms, the F1 generation showed GFP expression in the head, tail, muscles and excretory canal, whilst the F2 generation (and subsequent generations) expressed GFP in the

- 29 -

excretory canal with background expression in other cells. The background expression in the other cells is mosaic and depending from animal to animal the pharynx, the gut, muscle cells or the tail may show background GFP expression. The resulting strain was designated UG266 (bgEx34).

A smaller construct containing a 2.2 kb *SphI*-*MscI* fragment of UL6 in pPD95.79 was made and designated pGF2003. Following injection of pGF2003 into *C.elegans* worms, the F1 showed GFP expression in the head, tail, muscles and excretory canal, analogous to pGF2002. The expression of GFP in the F2 generation was not analysed but is expected to be analogous to that observed with pGF2002, i.e. strong expression in the excretory canal with background expression in other tissues.

Finally, a 0.9 kb *NruI*-*MscI* fragment of UL6 was cloned into pPD95.79 digested with *MscI* and *XbaI*, the later made blunt with Klenow polymerase. After injection of the resultant plasmid, designated pGF2001, into *C. elegans* no GFP expression could be observed.

These results and the fact that the excretory cell-specific promoter in the UL6 fragment was thought to be localized between the two genes, being approximately between the *NruI* and the *SphI* site, stimulated the inventors to analyse further deletion constructs. The aim was to determine the smallest fragment of UL6 that promotes the transcription and expression in the *C. elegans* excretory cell and excretory canal.

The 1.3 kb *NruI*-*SphI* insert of pGF2003 was excised and cloned into the vector pPD97.78. The

- 30 -

latter vector is analogous to pPD95.79, but contains a minimal *pes-10* promoter upstream of the GFP gene. This means that it is not necessary to make a fusion construct between the fragment to be tested for promoter activity and GFP. Vector pPD97.78 was
5 digested with *Sph*I and *Hind*II and the resulting plasmid was designated pGF2006. Transgenic *C. elegans* animals harboring this plasmid from the F2 and subsequent generations show strong GFP expression in
10 the excretory canal with only minor background expression in other cell types. The new *C. elegans* strain was designated UG267(bgEx35).

A 1.7 kb *Msc*I-*Sac*I fragment of pGF2003 was cloned in pPD95.79 and designated pGF2008. After injection of
15 this plasmid into *C. elegans* worms the F1 generation showed GFP expression in the excretory canal, gut, head, tail and spermatheca.

The region that promotes the expression of GFP in the excretory canal seemed to be located in the 1.3 kb
20 *Nru*I-*Sph*I insert of pGF2006. Further deletion analysis was therefore carried out to identify the minimal UL6 promoter in this fragment. Plasmid pGF2009 was generated by deleting a 0.5 kb *Bpi*I-*Hind*III fragment of pGF2006, leaving a 0.9 kb fragment. The *Hind*III
25 site is part of the multiple cloning site of the vector and adjacent to the *Sph*I site. Plasmid pGF2010 was generated by deleting a 0.9bp *Xba*I-*Bpi*I fragment of pGF2006, leaving a 0.5 kb fragment. The *Xba*I site is part of the multiple cloning site of the vector and
30 adjacent to the *Nru*I-*Hind*II fusion. The construction of this series of deletions will be further understood with reference to Figure 8 which shows a restriction map of the UL6 insert and Figure 9 which gives an

- 31 -

overview of the UL6 deletion constructs.

C. elegans injected with pGF2009 showed GFP expression in the excretory canal, with only minor background expression in the other cells, analogous to the expression results with pGF2006. The resulting *C. elegans* strain was designated UG271 (bgEX38). No GFP could be detected in the F1 generation after introduction of pGF2010 into *C. elegans*.

Two further deletions have been constructed to define the minimal promoter capable of directing transcription in the excretory canal. pGF2009 was digested with restriction enzyme *Bsa*I and made blunt with klenow polymerase. The resulting fragment was further digested with *Xba*I to give a 0.4 kb *Bsa*I-*Xba*I fragment and with *Hind*III to give a 0.5 kb *Hind*III-*Bsa*I fragment. These fragments were independently cloned in pPD97.78, using the *Hind*III-*Stu*I and *Stu*I-*Xba*I sites, respectively. The resulting vectors are designated pGF2011 and pGF2012. Injection of pGF2012 into *C. elegans* resulted in expression of GFP in the excretory canal.

Plasmid expression vectors suitable for expressing GFP or lacZ or any other reporter protein in the excretory canal of *C. elegans* can be integrated in the genome of the worm using standard technology. The vectors pGF2006, pGF2009 and pGF2012 are particularly suitable for this purpose.

Example 2

Methods of screening for new mutations.

To develop a screen for new mutants having a

- 32 -

"short canals" phenotype a *C. elegans* strain which contains any of the above plasmids that express GFP in the excretory canal can be used. To develop a screen for new mutants with a "ventral canals" phenotype a
5 worm strain with a stably integrated GFP-expressing plasmid is crossed with a *C. elegans* strain having the UNC phenotype, such as *C. elegans* strain MT152, *unc-53* (n152) (obtained from Dr Bob Horvitz MIT, Cambridge MA, USA). F2 worms with short canals (UNC phenotype)
10 are used for further screening.

General protocol for mutagenesis.

A few thousand adult worms of the desired genetic background are treated with a hypochlorite solution to
15 get a synchronised culture. This culture is then mutagenized using the EMS technique when the worms have reached the L4 stage (protocol for mutagenesis is as described in "Methods in Cell Biology, Vol 48 page 31-35"). As an alternative to EMS the UV-TMP technique
20 can be used. In this method the worms are contacted with tri-methyl-psoralen and then treated with UV radiation, as described in Methods in Cell Biology, Vol 48. *Caenorhabditis elegans*: Modern biological analysis of an organism. Eds H. F. Epstein and D. C.
25 Shakes, Academic Press. F2 worms are analysed for aberrant excretory canal phenotypes, which can easily be visualised due to the expression of GFP.

Observations of excretory canal morphology are
30 made for two subsequent generations following contact with mutagen. Dependent on the starting genetic background of the worms, the following defects can be observed in the progeny:- nematodes having shorter or

- 33 -

longer canals, nematodes having curved or extra
branched canals, nematodes having ventral or dorsal
canals, nematodes having more less than two canals,
nematodes having wrongly branched canals , nematodes
5 having vacuoles or cysts, nematodes with unusual
features in the excretory canal.

Example 3

10 **A method of detecting compounds that influence the
excretory canal phenotype.**

Worms stably expressing GFP in the excretory
canal can be used to detect and isolate compounds that
have effects on the morphology of the excretory canal
15 as follows:

Standard agar plates for use with *C. elegans* are
seeded with *E.coli* and left to full growth. Serial
dilutions of compound to be tested are then pipetted
20 onto the *E.coli* lawn and allowed to diffuse into the
lawn. One transgenic *C. elegans* expressing GFP in the
excretory canal (L4 stage) per dilution of compound is
put onto the bacterial lawn. The plates are incubated
at 21°C and visually screened for excretory canal
25 morphology at various time intervals and for two
generations to detect aberrant phenotypes. The
following defects in excretory canal morphology can be
observed; nematodes having shorter or longer canals,
nematodes having curved or extra branched canals,
30 nematodes having ventral or dorsal canals, nematodes
having more or less than two canals, nematodes having
wrongly branched canals, nematodes having vacuoles or
cysts, nematodes with unusual features in the

- 34 -

excretory canal.

Example 4

5 Use of the UL6 minimal promoter fragment in expression of a heterologous DNA in the excretory canal.

The above-described plasmids pGF2002, pGF2003, pGF2006, pGF2008, pGF2009 and pGF2012 have all been used to express GFP in the excretory cell and
10 excretory canal of *C.elegans*.

The same plasmids can be used to drive the transcription of other DNA fragments than the GFP encoding fragment by simply replacing this GFP encoding fragment with any other DNA fragment of
15 interest. Preferentially the vectors pGF2006, pGF2009 or pGF20012 are used for this purpose. As an alternative any of the isolated and analysed promoter fragments described above and shown in Figures 2-7 can be cloned upstream of the target DNA of interest in an
20 expression vector suitable for use in *C. elegans*.

Example 5

Compound screening assay.

25 The following method may be used to determine whether a compound is a modulator of growth cone steering, cell shape, cell motility, tumour formation, tumour vascularisation, metastasis, renal development, a pathway involved in kidney disease, development of
30 the excretory canal, cytoskeletal organisation or surface to cytoskeleton signalling:

Standard agar plates for use with *C. elegans* are

- 35 -

seeded with *E.coli* and left to full growth. Serial dilutions of compound to be tested are then pipetted onto the *E.coli* lawn and allowed to diffuse into the lawn. One transgenic *C. elegans* (L4 stage) per
5 dilution of compound is put onto the bacterial lawn. The transgenic *C. elegans* is one expressing both a protein involved in the regulation of growth cone steering, cell shape, cell motility, tumour formation, tumour vascularisation, metastasis, renal development,
10 a pathway involved in kidney disease, development of the excretory canal, cytoskeletal organisation or surface to cytoskeleton signalling and GFP in the excretory canal. The plates are incubated at 21°C and visually screened for excretory canal morphology at
15 various time intervals and for two generations to detect aberrant phenotypes.

The following abnormal excretory canal morphologies can be observed; nematodes having shorter or longer canals, nematodes having curved or extra
20 branched canals, nematodes having ventral or dorsal canals, nematodes having more or less than two canals, nematodes having wrongly branched canals nematodes having vacuoles or cysts, nematodes with unusual features in the excretory canal.

25

Example 6

Construction of a *C. elegans* strain stably expressing a reporter gene in the excretory canal.

30 Although expression of reporter genes in the excretory canal from extrachromosomal arrays is sufficient to perform most applications, stable

- 36 -

expression in *C. elegans* by integration would facilitate and improve most if not all the applications. Several *C. elegans* strains have been constructed, wherein the DNA encoding for the promoter (here designated as UL6) that drives the expression of the reporter gene, in this case GFP, is integrated the in the genome.

Method

A general method to integrate extrachromosomal DNA in to the genome of *C. elegans* has been described by Mello and Fire in Methods in Cell biology, Volume 48, *Caenorhabditis elegans*: Modern biological analysis of an organism, Chapter 19, 466-468. Here the construction of a particular example is described in detail:

A wild type N2 *C. elegans* strain was injected with pGF2006 (100ng/μl) and with pUC18DNA (100ng/μl) using standard methods, resulting in strain UG489 *bgEx176*. Although a relatively high transmission efficiency (up to 55%) was observed, which troubles later selections, a clean expression of GFP was observed in the excretory cell and the excretory canal.

9cm NGM plates with a population of late stage L4 worms were gamma-irradiated for 1 hour in order to obtain a total intensity of 30 gray. 6 x 35 animals were isolated after irradiation, incubated for 24 hours at 20°C and then transferred to fresh plates and further incubated at 25°C. 2 x 500 F1 animals were isolated after 48 hours and after 72 hours respectively, and further incubated at 25°C. 2-4 F2's per F1 were isolated and incubated at 15°C.

- 37 -

The progeny of these F2's was checked by selecting for plates with 75% or 100% of the progeny expressing GFP. Of these positive plates, 4 F3's were isolated and further incubated at 15°C. Again the progeny was
5 checked for the presence of 75% to 100% of the worms expressing GFP.

The resulting integrated lines were crossed out several times to confirm that no major translocations or unlinked mutations have occurred as a result of the
10 radiation treatment. This was done by mating wild-type *C. elegans* (N2) males with the hermaphrodites carrying the integration (the N2 strain can be obtained from CGC, University of Minnesota, USA). F1 males are used to cross to the mapping strains and to cross back to
15 N2 hermaphrodites (crossing out once every generation by always using GFP male progeny, except if the site of the integration is on X).

The integrations were mapped to the *C. elegans* chromosomes, by applying standard techniques well
20 known in the art. Mapping was performed using at least following strains: MSI dpy-5(e61)I;unc-4(e120)II; lon-1(e185)III and MT464 unc-5(e53)IV; dpy-11(e224)V; lon-2(e678)X.

25 **Result**

Of a total of 1000 F1 worms primary isolated, 650 lines were further examined. Out of these examined lines, eight integrated lines were retained, crossed out and mapped:
30 UG703 bgIs309 has strong hypodermal GFP expression, integrated on X;
UG704 bgIs310 has low hypodermal GFP expression, integrated on X;

- 38 -

UG705 bgIs311 has nearly no hypodermal GFP expression,
integrated on X;
UG706 bgIs312 has nearly no hypodermal GFP expression,
integrated on I;
5 UG707 bgIs313 has strong hypodermal GFP expression,
integrated on IV;
UG708 bgIs314 has nearly no hypodermal GFP expression,
integrated on X;
UG709 bgIs315 has low hypodermal GFP expression,
10 integrated on X;
UG710 bgIs316 has low hypodermal GFP expression,
integrated on V.

Conclusion

15 Of the eight independent integrated lines obtained, at
least five are of good quality (bgIs311, bgIs312,
bgIs314, bgIs315 and bgIs316).

20 Example 7

Construction of improved vectors to express genes in the *C. elegans* excretory cell.

The plasmid expression vectors pGF2006 and
pGF2009 for expression of the reporter gene GFP both
25 contain a nuclear localisation signal (NLS) in fusion
with the reporter gene. Expression of the fusion
protein (NLS-GFP) results in the translocation of a
part of the total amount of the protein expressed to
the nucleus of the excretory cell. Although the
30 expression of the fusion protein is high enough to
visualise the fusion protein in the whole excretory
cell including the excretory canal, the present
inventors decided to delete this NLS part. Expression

from the UL6 promoter would hence no more result in translocation in the nucleus of the expressed gene, and provide for a more equal localisation of the expressed gene.

5 To test this hypothesis the NLS signal was deleted from the plasmids pGF2006 and pG2009. This was easily done by deletion of a KpnI restriction fragment resulting in the plasmids pGF2013 and pGF2014 respectively.

10

SEQUENCE LISTING

SEQ ID NO: 1 is the nucleotide sequence of the
insert of pUL6#64A1, illustrated in
Figure 1.

SEQ ID NO: 2 is the nucleotide sequence of the insert of pGF2002, illustrated in Figure 2.

20

SEQ ID NO: 3 is the nucleotide sequence of the insert of pGF2003, illustrated in Figure 3.

SEQ ID NO: 4 is the nucleotide sequence of the
25 insert of pGF2006, illustrated in
Figure 4.

SEQ ID NO: 5 is the nucleotide sequence of the
insert of pGF2008, illustrated in
30 Figure 5.

SEQ ID NO: 6 is nucleotide sequence of the insert of pGF2009, illustrated in Figure 6.

- 40 -

SEQ ID NO: 7 is the nucleotide sequence of the
insert of pGF2012, illustrated in
Figure 7.

5 SEQ ID NO: 8 is the complete nucleotide sequence of
plasmid pGF2006.

SEQ ID NO: 9 is the complete nucleotide sequence of
plasmid pGF2009.

10

SEQ ID NO: 10 is the complete nucleotide sequence of
plasmid pGF2013.

15

SEQ ID NO: 11 is the complete nucleotide sequence of
plasmid pGF2014.

The contents of any published patents, patent
applications and scientific publications referred to
20 herein are to be incorporated herein by reference.

Claims:

1. A DNA fragment which is capable of functioning as a promoter directing gene expression in the excretory cell of *C. elegans*, which DNA fragment comprises the sequence of nucleotides set forth in Figure 2 or a fragment thereof in the absence of any other sequence of consecutive nucleotides from the *C. elegans* genome.

10

2. A DNA fragment which is capable of functioning as a promoter directing gene expression in the excretory cell of *C. elegans*, which DNA fragment comprises the sequence of nucleotides set forth in Figure 3 or a fragment thereof in the absence of any other sequence of consecutive nucleotides from the *C. elegans* genome.

3. A DNA fragment which is capable of
functioning as a promoter directing gene expression in
the excretory cell of *C. elegans*, which DNA fragment
comprises the sequence of nucleotides set forth in
Figure 4 or a fragment thereof in the absence of any
other sequence of consecutive nucleotides from the *C.*
elegans genome.

4. A DNA fragment which is capable of functioning as a promoter directing gene expression in the excretory cell of *C. elegans*, which DNA fragment comprises the sequence of nucleotides set forth in Figure 5 or a fragment thereof in the absence of any other sequence of consecutive nucleotides from the *C. elegans* genome.

30

- 42 -

5. A DNA fragment which is capable of functioning as a promoter directing gene expression in the excretory cell of *C. elegans*, which DNA fragment comprises the sequence of nucleotides set forth in Figure 6 or a fragment thereof in the absence of any other sequence of consecutive nucleotides from the *C. elegans* genome.

6. A DNA fragment which is capable of functioning as a promoter directing gene expression in the excretory cell of *C. elegans*, which DNA fragment comprises the sequence of nucleotides set forth in Figure 7 or a fragment thereof in the absence of any other sequence of consecutive nucleotides from the *C. elegans* genome.

7. An expression vector which is suitable for directing tissue-specific expression of a heterologous DNA fragment in the excretory cell of *C. elegans*, said vector comprising a promoter comprising a DNA fragment as claimed in any one of claims 1 to 6 positioned to direct expression of the said heterologous DNA fragment.

8. An expression vector as claimed in claim 7 wherein the heterologous DNA fragment is a reporter gene.

9. An expression vector as claimed in claim 8 wherein the reporter gene encodes green fluorescent protein, β -galactosidase, β -lactamase, luciferase, acetohydroxyacid synthase, alkaline phosphatase, β -glucuronidase, chloramphenicol acetyltransferase,

- 43 -

horseradish peroxidase, nopaline synthase or octapine synthase.

10. A host cell transformed or transfected with
5 an expression vector as claimed in any one of claims 7 to 9.

11. A transgenic *C. elegans* containing a
transgene comprising a promoter which is capable of
10 directing tissue-specific gene expression in the excretory cell of *C. elegans* operatively linked to a protein-encoding DNA fragment.

12. A transgenic *C. elegans* as claimed in claim
15 11 wherein the promoter comprises a DNA fragment as claimed in any one of claims 1 to 6.

13. A transgenic *C. elegans* as claimed in claim
11 or claim 12 wherein the protein-encoding DNA
20 fragment comprises a reporter gene encoding green fluorescent protein, β -galactosidase, β -lactamase, luciferase, acetohydroxyacid synthase, alkaline phosphatase, β -glucuronidase, chloramphenicol acetyltransferase, horseradish peroxidase, nopaline
25 synthase or octapine synthase.

14. A transgenic *C. elegans* as claimed in any
one of claims 11 to 13 wherein said transgene is
stably integrated into a chromosome of the said *C.*
30 *elegans*.

15. A transgenic *C. elegans* as claimed in claim
11 or claim 12 which further comprises a second

- 44 -

transgene, which transgene comprises a promoter suitable for directing tissue-specific gene expression in the excretory cell of *C. elegans* operatively linked to a reporter gene.

5

16. A transgenic *C. elegans* as claimed in claim 15 wherein the promoter comprises a DNA fragment as claimed in any one of claims 1 to 6.

10

17. A transgenic *C. elegans* as claimed in claim 15 or claim 16 wherein the reporter gene encodes green fluorescent protein, β -galactosidase, β -lactamase, luciferase, acetohydroxyacid synthase, alkaline phosphatase, β -glucuronidase, chloramphenicol acetyltransferase, horseradish peroxidase, nopaline synthase or octapine synthase.

15

18. A transgenic *C. elegans* as claimed in any one of claims 15 to 17 wherein one or both of the transgenes is integrated into the chromosome of the said *C. elegans*.

20

19. A method of identifying a mutation in a gene involved in growth cone steering, cell motility, cell shape, tumour formation, metastasis, vascularisation of tumours, renal development, kidney disease, the development of the excretory canal, cytoskeletal organisation or surface to cytoskeleton signalling, which method comprises contacting a transgenic *C. elegans* which expresses a reporter gene in the excretory canal with a mutagen and screening for phenotypic changes in the excretory canal.

25

30

- 45 -

20. A method as claimed in claim 19 wherein the mutagen is EMS, UV-TMP or X-rays.

21. A method as claimed in claim 19 or claim 20
5 wherein the transgenic *C. elegans* contains a transgene comprising a promoter which is suitable for directing tissue-specific gene expression in the excretory canal of *C. elegans* operatively linked to a reporter gene.

10 22. A method as claimed in claim 21 wherein the promoter comprises a DNA fragment as claimed in any one of claims 1 to 6.

15 23. A method as claimed in any one of claims 19 to 22 wherein the transgenic *C. elegans* is a wild-type strain or a selected mutant strain.

20 24. A method as claimed in any one of claims 19 to 23 wherein the reporter gene encodes green fluorescent protein, β -galactosidase, β -lactamase, luciferase, acetohydroxyacid synthase, alkaline phosphatase, β -glucuronidase, chloramphenicol acetyltransferase, horseradish peroxidase, nopaline synthase or octapine synthase.

25

25. A method of determining whether a compound is an inhibitor or an enhancer of the regulation of growth cone steering, cell motility, cell shape, tumour formation, metastasis, vascularisation of
30 tumours, renal development, pathways involved in kidney disease, development of the excretory canal, cytoskeletal organisation or surface to cytoskeleton signalling, which method comprises contacting a sample

- 46 -

of the compound with a transgenic *C. elegans* which expresses a reporter gene in the excretory canal and screening for phenotypic changes in the excretory canal.

5

26. A method as claimed in claim 25 wherein said transgenic *C. elegans* is a wild-type strain or a selected mutant strain.

10

27. A method as claimed in claim 25 or claim 26 wherein the transgenic *C. elegans* contains a transgene comprising a promoter which is suitable for directing tissue-specific gene expression in the excretory canal of *C. elegans* operatively linked to a reporter gene.

15

28. A method as claimed in claim 27 wherein the promoter comprises a DNA fragment as claimed in any one of claims 1 to 6.

20

29. A method as claimed in any one of claims 25 to 28 wherein the reporter gene encodes green fluorescent protein, β -galactosidase, β -lactamase, luciferase, acetohydroxyacid synthase, alkaline phosphatase, β -glucuronidase, chloramphenicol acetyltransferase, horseradish peroxidase, nopaline synthase or octapine synthase.

25

30. A compound which is identifiable as an inhibitor or an enhancer of the regulation of growth cone steering, cell motility, cell shape, tumour formation, metastasis, vascularisation of tumours, renal development, pathways involved in kidney disease, development of the excretory canal,

30

- 47 -

cytoskeletal organisation or surface to cytoskeleton signalling using the method of any one of claims 25 to 29.

5 31. A compound as claimed in claim 30 for use as a medicament for promoting neuronal regeneration, re-vascularisation or wound healing.

10 32. A compound as claimed in claim 30 for use as a medicament for the treatment of metastasis, tumour formation, tumour vascularisation, chronic neuro-degeneration, kidney disease, kidney cyst formation, polycystic kidney diseases, cell migration diseases or immunological diseases.

15 33. Use of a compound as claimed in claim 30 in the manufacture of a medicament for promoting neuronal regeneration, re-vascularisation or wound healing.

20 34. Use of a compound as claimed in claim 30 in the manufacture of a medicament for the treatment of metastasis, kidney disease, kidney cyst formation, polycystic kidney diseases, cell migration diseases or immunological diseases.

25 35. A pharmaceutical composition comprising a compound as claimed in any one of claims 30 to 34 and a pharmaceutically acceptable carrier, diluent or excipient therefor.

30 36. A method of determining the function of the protein encoded by a DNA fragment, which method comprises the steps of;

- 48 -

(a) providing a transgenic *C. elegans* containing a transgene comprising a promoter suitable for directing tissue-specific gene expression in the *C. elegans* excretory cell operatively linked to the said DNA fragment; and

(b) observing any phenotypic changes in the excretory canal of the transgenic *C. elegans*.

37. A method as claimed in claim 35 wherein the promoter comprises a DNA fragment as claimed in any one of claims 1 to 6.

38. A method as claimed in claim 36 or claim 37 wherein the transgenic *C. elegans* further comprises a second transgene comprising a promoter suitable for directing tissue-specific gene expression in the excretory cell of *C. elegans* operatively linked to a reporter gene.

39. A method as claimed in claim 38 wherein the promoter of the second transgene comprises a DNA fragment as claimed in any one of claims 1 to 6.

40. A method as claimed in claim 38 or claim 39 wherein the reporter gene encodes green fluorescent protein, β -galactosidase, β -lactamase, luciferase, acetohydroxyacid synthase, alkaline phosphatase, β -glucuronidase, chloramphenicol acetyltransferase, horseradish peroxidase, nopaline synthase or octopine synthase.

41. A method of determining whether a compound is a modulator of growth cone steering, cell shape,

- 49 -

tumour formation, metastasis, vascularisation of tumours, cell motility, renal development, a pathway involved in kidney disease, development of the excretory canal, cytoskeletal organisation or surface to cytoskeleton signalling, which method comprises the steps of;

(a) contacting a sample of the compound with a transgenic *C. elegans* expressing a DNA fragment encoding a protein involved in the regulation of growth cone steering, cell shape, cell motility, renal development or a pathway involved in kidney disease, which transgenic *C. elegans* contains a transgene comprising a promoter which is suitable for directing tissue-specific gene expression in the *C. elegans* excretory cell operatively linked to the said DNA fragment; and

(b) screening for phenotypic changes in the excretory canal.

42. A method as claimed in claim 41, wherein said transgenic *C. elegans* is a wild-type strain or a selected mutant strain.

43. A method as claimed in claim 41 or claim 42 wherein the promoter comprises a DNA fragment as claimed in any one of claims 1 to 6.

44. A method as claimed in any one of claims 41 to 43 wherein the transgenic *C. elegans* further contains a second transgene comprising a promoter suitable for directing tissue-specific gene expression in the excretory cell of *C. elegans* operatively linked to a reporter gene.

- 50 -

45. A method as claimed in claim 44 wherein the promoter of the second transgene comprises a DNA fragment as claimed in any one of claims 1 to 6.

5 46. A method as claimed in claim 44 or claim 45 wherein the reporter gene encodes green fluorescent protein, β -galactosidase, β -lactamase, luciferase, acetohydroxyacid synthase, alkaline phosphatase, β -glucuronidase, chloramphenicol acetyltransferase,
10 horseradish peroxidase, nopaline synthase or octapine synthase.

 47. A method as claimed in any one of claims 41 to 46 wherein the transgenic *C. elegans* has an
15 abnormal excretory canal phenotype.

 48. A method as claimed in any one of claims 41 to 47 wherein the DNA fragment expressed in the excretory cell of the transgenic *C. elegans* rescues an
20 abnormal excretory canal phenotype which is present in the genetic background of the transgenic *C. elegans*.

 49. A compound which is identifiable as a modulator of growth cone steering, cell shape, cell
25 motility, tumour formation, metastasis, vascularisation of tumours, renal development, a pathway involved in kidney disease, development of the excretory canal, cytoskeletal organisation or surface to cytoskeleton signalling using the method of any one
30 of claims 41 to 49.

 50. A compound as claimed in claim 49 for use as a medicament for promoting neuronal regeneration, re-

- 51 -

vascularisation or wound healing.

51. A compound as claimed in claim 49 for use as a medicament for the treatment of chronic neuro-
5 degenerative diseases, metastasis, tumour formation, tumour vascularisation, kidney diseases, kidney cyst formation, polycystic kidney diseases, cell migration diseases or immunological diseases.

10 52. Use of a compound as claimed in claim 49 in the manufacture of a medicament for promoting neuronal regeneration, re-vascularisation or wound healing.

15 53. Use of a compound as claimed in claim 49 in the manufacture of a medicament for the treatment of chronic neuro-degenerative diseases, metastasis, tumour formation, tumour vascularisation, kidney diseases, kidney cyst formation, polycystic kidney diseases, cell migration diseases or immunological
20 diseases.

54. A pharmaceutical composition comprising a compound as claimed in any one of claims 49 to 53 and a pharmaceutically acceptable carrier, diluent or
25 excipient therefor.

55. A method of identifying further components of a biochemical pathway on which a compound identifiable as a modulator of growth cone steering,
30 cell shape, cell motility, renal development, a pathway involved in kidney disease, development of the excretory canal, cytoskeletal organisation or surface to cytoskeleton signalling may act, which method

- 52 -

comprises the steps of:

(a) providing a transgenic *C. elegans* which expresses a reporter gene in the excretory canal;

5 (b) contacting the transgenic *C. elegans* with a mutagen;

(c) contacting the mutated *C. elegans* with a compound which is identifiable as a modulator of growth cone steering, cell shape, cell motility, renal development, a pathway involved in kidney disease,
10 development of the excretory canal, cytoskeletal organisation or surface to cytoskeleton signalling;
and

(d) screening for phenotypic changes in the excretory canal.

15

56. A method as claimed in claim 55 wherein the mutagen is EMS, UV-TMP or X-rays.

57. A method as claimed in claim 55 or claim 56
20 wherein the transgenic *C. elegans* contains a transgene comprising a promoter which is suitable for directing tissue-specific gene expression in the excretory canal of *C. elegans* operatively linked to a reporter gene.

25 58. A method as claimed in claim 57 wherein the promoter comprises a DNA fragment as claimed in any one of claims 1 to 6.

59. A method as claimed in any one of claims 55
30 to 58 wherein the reporter gene encodes green fluorescent protein, β -galactosidase, β -lactamase or luciferase.

FIG. 1.

gatcgcctcgaataaaaaattttataatgtcagattacggttttagatccaaaaaa
attaggctgaatccggtttttgattctctagaattttgttttagccaaaaaacatcg
cattgttctctctgctgcttatttaaagtgcgcaattaacaaatattcaaaaattct
tatttaaaaaaccataggggggttccctaactattgcgaaattctctttctctctc
ggacattactatgagagattgaaccaattgaagagacgcagtgctaggaagggaa
tgcgccctggcgccacatattggtctaagtctctctgaatcgtgtgctctctacat
tggttcgactctgcgagtgcttttgagcggcgccgagaccgaaataagagagaaa
acagagggaaaaaagataggggaacattaattgatgaaaaaagaaacagatgacgt
ggcaattcttctttccggagtttttttttatttgaagtggggttattcaagtaat
gtagcaaaatgtatttaaatacatttgtgacgtcacaaatgtatttaaatacatg
tttttatatacttgaataaggttgtgacgtaatttttctacactttttaattttc
cgacactacttgaataaccccaaaagtgtacggtttcttttttcaaaacacgatt
gcaaccaaaaggcgccggtttttgaatttttcaaaaatcggaagatttaaattt
tcgcttttttatttttattacttgataaaaaattgaattttatgggtgaaatttcaa
aatttgttttggcttttcagttgtccgaaaaatcgaaaaatgtttgtttttcgatt
cccactgaaaaaatcgaattttttcgatgaaacctttgattgcaatgttagaaaa
atttctaaaaaatttaacaagaacactggaacaacacaaaatcacataatcaaca
ccggaccattggtaccgtataaaaaagaaaaaaaatcgaattaattgtaatagtt
ctgaaaaatctttcataaatacccattttgaatataattatttcttaataaatata
taacttacagagaaaaaaatttcggggacgttaaggctcatagaagaacacacga
aaattaggttaagtttaaggctgaaggcatttggcctactctacacgtagcagggaa
atgatgtaaaatggaaaagagagaaaaatatataaacatatgaatgtgcaaggat
tcggtatataaattatatatgaagggtccttattaaggaatttgttttctctgtat
gtacaatctagaaaagagcagaagggttattgaaagggtggagtagagtcaaaat
tgtttttttacttttaaattacagaaaaatggacactaatgacggaatataactata
aacatttttttctaaattttgaaaaatgattttttcaattttgcacctctcaaatt
aagagcactaaagaaaattacagaaaaaccagtaaaatttgagaatttgaaacgg
atatctccaaaaagtagtcgaagtctaagaatcaaaaaagaaattaaaatttttt
ttagttacggtatattttctgtcattttaaagcagttgactccactccaccttaa
aacaatataaattatctaattgaggcctagcagaatatcggttctctcaaatacggga
ttatagggtgtcatatcgcttttctccattcgctccaaacgattttggcgacaa
ccagatgaaaagcgacgacgacacaaaaaacgaacaccagaatgttgtcattcc
tccagttgattctccagatcgatgcatttcaataacttcttcttcgtgtgcccgt
gctgcaccaacttccgacacgtgaatggcatttcttaggagttcaagatgtcgtc
gagatggagcaatgtagagaagagatgttactgggtttcgtgacttttactggtaa
ttgaactggagtcctttccaagtgaactgggctatcgaggcatccaccatcacaa
acaaataattgactagaagatgaaacctgaaagaaatagtttagagtatttccag
gataaataatttttaaaatttaaataaaaagacgtgtttcagctcttttgacaagga
tttttattcattttttgagaagtcctactatgaaattcggggcccagttcagtcata
tttaagcaacaagaaaccacaaactatgctacagttgtaaaacattctataaact
attaccataaaaataaggacgattttccgatatccaattttatatccactcgaatag
taccaacaacacccatattcgaataaaactggaattccatgaaaattcatttcacg
atgcaaatcatgagcatccaatgctacttccaaatgagaatcgtgataactcgat
gcagtcacagattgaacaaatgcttgtgatgctccaattgctccagtgctcaaaa
gattgttgcaacctcggttttggaaagtgagcatccatgtattcacgagctcaaa
aagttttgtcggagtatcccagtccttgccacagacgtgatgggaagagaagattt
gaatgaccagtgtagcagagctcgcttttccgcgccaaattttccagaggctata
aacattctatttttaattaggtaataatattcaattacctttacagatttccaaagt
agcccatgctcaattgttggagcccgtaaaattgctcattaaaatgtattttgtt
gcattaattctatctccgtttaatgtattttgggtgctttcgaattcgcaacactga
atcctgttaaatgcattcttgcctccatccgtcaaaatgctccgaatccaaagtttt

FIG. 1.(CONTINUED)

cgagtcttttctgattggaaaaatcggatagcatcccgagaagcttcttcactt
gttggcttcgacgcaatcgagcatgtaaaccgaaatgactcttctctgaaataaa
ctgaacaaattaattaatttttttaattgttttaaatataccttttctgtgaaact
gaaacttttccagggtgcatcagtacaaacaacagcatagctctagcaactt
gccgctgatgaaccgtcagctcgcctaatttttgaatgagcccaccagttggaact
aagacgagacaagttgagtcggaccgttgatttctgaaattaaatgttcttttg
taattttaatgcatcaatttcaataaaaattacagttggattggaaacttgatca
attgaacaaaatagtttgagcaggttcgatgagcccatatctgtgacgtcacaca
tgcacattcgccatcctgaaacattaattgtccaataaatagtttattatgtcgtg
gctaaccattgaaaaacatcgcagaattttgaaaatgccatttttgaaatccgtc
atttttgtaacctcaacaggttggttgcgtcaatttcaatgttgagcccggtat
aaaaattggtttccacatcgagaactcgggtttttctcctgaactcgtaattgtcc
gaatgaatcaacaccaatataatccatttcccacaaactctacgctttgattgtta
attcctgtatcgaaactcctaaaaaaaagtaaatatgaataattactaattattcg
actcacatcaatctctttttcccataatttttaacttttccgctgaacattgatgc
tgttttgtgaatgatgtgccggaagacgttctcgcgagaagatgaaaatgacga
ataaaacaactccaataaattatcgctaattctcgttttggtatcgtcgaaccat
acgcattttactttcgaaatccatattttctattcgacgcacgaaagtgcgccac
gttattgcggtcgtcattttggaggggaaatccgcctgaaaaatcaattgtttgt
attgtgaaatttcgaagaggcataaaaacaagaaaacggacatgaaagcgcgttgc
atgcaagggttagttgcctgtttaagcattatccccgcgatgtagcttgttcggcac
cgttaaaaatgctgagtaatcagctttttagaattttaaataattaaacttttaa
attgcaacaaacatcgacaaatattcaagaggcgaatgatatcggaatttcgat
tgaaacgaaactgttttgaaattcaaaaagtattttcaaagtattgtccgcaagg
cacatcacgcaaacttgagaatctaccgtatcccatacatttttatagtttttc
cctcagtttttaaaattaaaaacgctgaaaaagcgattaaattttatttaaattgca
tcgttcgaataaaaataaagtttatcttttgataaaaacatgagtttcttggag
aaaagtagggttttcgcctttcaaaaaattatttcgtgcaggatgctattttcgt
ggcgaaaccatactcaagagctcatgcgtcttcttgattactgtagatgtttgg
caacttatttttacataaaaacgttttcattcattatttccatcattcatttatc
tttctgtgttttttagtttagtttttagctagtttttttctaaattcctaactttaaa
aaatctggaaaagaaattaaaaaattttgtccctattattttattttattactgg
aaaatcttcaaacaggaaaaccaccgcttttgcttattgctgtatttatgaaa
aaaaaaacaatcaatattggtcaagtaataagaaaaaattaacgaatctctatc
tgacaccagatgcgaccctctattccaacttctctgttcattctgctgcttcttttg
tttaaccagataaatctccctcggggaaaaccgtcaaaaaaggcaactaaatg
caaacacgctctatagacaaaatgtgtttgggtctcgtcacgaatggtgagagaga
attggcctccgcgcgagagatcgcttgattattggcctccagtgggcaatgtcgg
ggaacacaaactattgatgagaggtatcgacgaaaaatcaaCaatgaccaactt
tttgttacagttttgttataaatatgagttttggatattccattgcgtatttttc
tttctactttcaaaaaatctgctccaacctttaatggcttttctgtcttgtca
aaatctggatttttgaatatataatttttaaaaccatcaaattcagcgaaatgaa
atcatgtaatacaattttttattttttccgactgttgtgtattccatcaaactat
tcaaaaaatcaataataatgatttttttttccatttttcgcgattttttattttt
gtcgtctgaaaacctttttactaataaaaataatttacagggaaaaccactaacga
ctgtagccatgggaatcagcgacaacgacgttcagaagcagctccgccacatgat
ggctttcattgagcaagaggccaatgagaaggctgaggagatcgatgctaaagcc
gaggaagaattcaacattgagaaaagtaaggaattaaaacatttactcctttaaaa
ctatactaaaaatctcttctaaaaaacggaaaaccttgaaattatgaattcattca
aattgtttcagggacgtcttgttcaacaacaacgtcaaaagattatggaattctt
cgagaagaaggagaaacaagtcgagcttcaacgcaaaattcaagcctccaactct

FIG. 1. (CONTINUED)

ctcaacgctggacgtcttcgcttgcttgaaggtgagagaaaacgtttctcaacatt
ttcaaaaacattaatcgcccttaaaattgaaaaccagttctgaatcggacacattt
gaattaaaaacatatattttcaggctcgtgaagaccacatcggagccgtactcgacg
aggctcgctcgaatctctcccgtatttccggagatgctgctcgttatccagctat
tttgaagggaacttgatcatgcaaggacttcttcaattgctcgaaaaggaagtcgtc
cttcgcttgccgtgagaaggatcttcgtcttggttgagcaacttttgccagagtgcc
ttgacggacttcaaaaaggagtggggaagcaccaccaaggctcgttctcgataaaca
aaacttcttgccatcggagtctgctggaggagttgaactttctgctcgtgctgga
aagatccccgggattggcca

FIG. 2.

aagcttcttccacttgttggcttcgacgcaatcgagcatgtaaaccgaaatgactc
ttctctgaaataaactgaacaaattaattaatttttttaatgttttaaatataacc
ttttctgtgaaactgaaactttttccaggtgcatcagtacaaacaacagcatacga
tgctctagcaacttgccgctgatgaaccgtcagctcgctaatttttgaatgagcc
caccagttggaactaagacgagacaagttgagtcggaccgttgatttctgaaatt
aaatgttctttttgtaattttaatgcatcaatttcaataaaaattacagttggatt
ggaaacttgatatcaattgaacaaaatagtttggacgagttcgatgagcccatatc
tgtgacgtcacacatgcacattcgccatcctgaaacattaatgtccaataaatag
tttattatgtcgtggctaaccattgaaaaacatcgagaattttgaaaatgccat
ttttgaaatccgtcattttttgtaacctcaacaggttgttgtccgtcaatttcaat
gttgagcccgggataaaaaattggtttccacatcgagaactcggtttttctcctga
actcgtaattgtccgaatgaatcaacaccaatataatccatttcccacaaactcta
cgctttgattgttaattcctgtatcgaaactcctaaaaaaaagttaaatatgaataa
ttactaattattcgactcacatcaatctctttttcccataattttaacttttccg
ctgaacattgatgctgttttgtgaatgatgtgccggaagacgttcctcgcgagaa
gatgaaaatgacgaataaaacaactccaataattatcgctaattctcgttttgga
tatcgtcgaaaccatacgcattttactttcgaaatccatattttctattcgacgca
cgaaagtgcgccacgttatttgccgctcatttttgagggggaaatccgcctgaaa
aatcaattgtttgtattgtgaaatttcgaagaggcataaaacaagaaaacggaca
tgaaagcgcgttgcatgcaaggttagttgcctgtttaagcattatccccgcattgt
agcttggtcggcacccgttaaaaaatgctgagtaatcagctttttagaattttaa
attaaactttttaaattgcaacaaacatcgacaaatattcaagaggcgaatgata
tcgggaattttcgattgaaacgaaactgttttgaaattcaaaaagtattttcaaag
tattgtccgcaaggcacatcacgcaaacttgagaatctaccgtatcccatacat
ttttatagtttttccctcagttttttaaattaaaaacgctgaaaaagcgattaaa
tttatttaaattgcatcgttcgaataaaataaagtttatcttttgataaaaacatg
agtttccctttggagaaaagtagggatttgcctttcaaaaaattatttctgtgcag
gatgctattttcgtggcgaaaccataactcaagagctcatgcgtcttcttgatta
ctgtagatgtttggcaacttatttttacataaaaaacgttttcattcattatttcc
atcattcatttatctttctgtgttttttagtttagtttttagctagttttttctaaa
ttcctaactttaaaaaatctggaaaagaaaattaaaaaattttgtccctattatt
tattttattactggaaaatcttcaaacaggaaaaccaccgcgttttgcttattg
ctgtattttatgaaaaaaaaaacaatcaatattggtcaagtaaataagaaaaaatt
aacgaatctctatctgacaccagatgcgaccctctattccacttctctgttcac
tgctgcttcttttgtttaaccagataaatctccctcggggaaaaccgtcaaaaaa
aggcaaaactaaatgcaaacacgctctatagacaaaatgtgtttggtctcgtcacg
aatggtgagagagaattggcctccgcgcgagagatcgcttgattattggcctcca
gtgggcaatgtcggggaaaaccacaaactattgatgagaggtatcgacgaaaaatca
acaatgaccaactttttgttacagttttgttataaatatgagttttggatattcc
attgcgtatttttcttttctactttcaaaaaatctgctccaacctttaatggctt
ttcctgtcttgtcaaaaatctggatttttgaatatataattttttaaaccatcaaa
ttccatcaaactattcaaaaaatcaatataatgatttttttttcatttttcgcga
ttttttattattttgtcgtctgaaaaccttttactaataaaaataatttacaggg
aaaaccactaacgactgtagccatgggaatcagcgacaacgacgttcagaagcag
ctccgccacatgatggcttccattgagcaagaggccaatgagaaggctgaggaga
tcgatgctaagccgaggaagaattcaacattgagaaaagtaaggaattaaaacat
ttactccttttaaaactatactaaaatctcttctaaaaaacggaaaaccttgaat
tatgaattcattcaaatgttttcagggacgtcttggttcaacaacaacgtcaaaaag
attatggaattcttcgagaagaaggagaaacaagtcgagcttcaacgcaaaaattc
aagcctccaactctctcaacgctggacgtcttcgcttgcttgaaggtgagagaaaa

FIG. 2. (CONTINUED)

cgtttctcaacattttcaaaaacattaatcgcccttaaaattgaaaaccagttctg
aatcggaacacatttgaattaaaaacatattttcagggtcgtgaagaccacatcgg
agccgtactcgacgaggctcgctcgaatctctcccgtatttccggagatgctgct
cgttatccagctattttgaagggacttgctcatgcaaggacttcttcaattgctcg
aaaaggaagtcgtccttcggttgccgtgagaaggatcttcgtcttggttgagcaact
tttgccagagtgccctgacggacttcaaaaggagtggggaagcaccaccaaggctc
gttctcgcataaaacaaaacttcttgccatcggagtctgctggaggagtgaacttt
ctgctcgtgctggaagatccccgggattggcca

FIG. 3.

gcatgcaagggttagttgcctgtttaagcattatccccgcagtagcttggttcggc
accgttaaaaatgctgagtaatcagctttttagaattttaaataattaaactttta
aaattgcaacaaacatcgacaaatattcaagaggcgaatgatatcgggaatttcg
attgaaacgaaactgttttgaaattcaaaaagtattttcaaagtattgtccgcaa
ggcacatcacgcaaacttgagaatctaccgtatcccatacatttttatagtttt
tccctcagttttttaaattaaaaacgctgaaaaagcgattaaattttatttaaag
catcgttcgaataaaaataaagttttatcttttgataaaaacatgagtttcctttgg
agaaaagttagggatttcgcctttcaaaaaattatttcgtgcaggatgctattttc
gtggcgaaacccatactcaagagctcatgctgcttcttgattactgtagatgttt
ggcaacttatttttacataaaaacgttttcattcattatttccatcattcattta
tctttctgtgttttttagtttagtttttagctagtttttttctaaattcctaacttta
aaaaatctggaaaagaaaattaaaaaattttgtccctattattttattttattact
ggaaaatcttcaaacaggaaaacccaccgcgttttgcttattgctgtattttatga
aaaaaaaaaacaatcaatattgggtcaagtaataagaaaaaattaacgaatctcta
tctgacaccagatgcgacctctattccacttctctgttcacctgctgcttcttt
tgtttaaccagataaatctccctcggggaaaaccgtcaaaaaaaggcaaaactaaa
tgcaaacacgctctatagacaaaatgtgtttgggtctcgtcacgaatgggtgagaga
gaattggcctccgcgcgagagatcgcttgattattggcctccagtgggcaatgtc
ggggaaaaccaaactattgatgagaggtatcgacgaaaaatcaacaatgaccaac
tttttggttacagttttgttataaatatgagttttggatattccattgcgtatttt
tcttttctactttcaaaaaatctgctccaacctttaatggcttttctgtcttggt
caaaatctggatttttgaaatatataattttttaaaccatcaaattcagcgaaatg
aatcatgtaatacaattttttattttttccgactgttggtgattccatcaaact
attcaaaaaatcaatataatgattttttttcatttttccgcattttttattatt
ttgtcgtctgaaaacctttttactaataaaaataatttacagggaaaaccactaac
gactgtagccatgggaatcagcgacaacgacgttcagaagcagctccgccacatg
atggctttcattgagcaagaggccaatgagaaggctgaggagatcgatgctaaag
ccgaggaagaattcaacattgagaaagtaaggaattaaaaacatttactccttttaa
aactatactaaaaatctcttctaaaaaacggaaaaccttgaaattatgaattcatt
caaattgtttcagggacgtcttggttcaacaacaacgtcaaaaagattatggaattc
ttcgagaagaaggagaacaagtgcagcttcaacgcaaaattcaagcctccaact
ctctcaacgctggacgtcttcgttgcttgaaaggtgagagaaaacgtttctcaaca
ttttcaaaaacattaatcgcccttaaaattgaaaaccagttctgaatcggacacat
ttgaattaaaaacaatttttcagggtcgtgaagaccacatcggagccgtactcga
cgaggctcgtcgaatctctcccgtatttccggagatgctgctcgttatccagct
attttgaagggacttgctcatgcaaggacttcttcaattgctcgaaaaggaagtcg
tccttcgttgccgtgagaaggatcttcgtcttggttgagcaacttttgccagagtg
ccttgacggacttcaaaaggagtggggaagcaccaccaaggctcgttctcgataaa
caaaacttcttgccatcggagtctgctggaggagtgaactttctgctcgtgctg
gaaagatccccgggattggcca

FIG. 4.

gcatgcaaggttagttgcctgtttaagcattatccccgcatgtagcttggttcggc
accgttaaaaaatgctgagtaatcagcttttttagaattttaaataattaaactttta
aaattgcaacaaacatcgacaaatattcaagaggcgaatgatatcggaatttcg
attgaaacgaaactgttttgaaattcaaaaagtattttcaaagtattgtccgcaa
ggcacatcacgcaaaacttgacagaatctaccgtatcccatacatttttatagttt
tccctcagttttttaaataaaaaacgctgaaaaagcgattaaatttatttaaag
catcgttcgaataaaaataaagtattatcttttgataaaaaacatgagtttcctttgg
agaaaagtagggatttcgcctttcaaaaaattatttcgtgcaggatgctattttc
gtggcgaaaaccatactcaagagctcatgcgtcttcttgattactgtagatgttt
ggcaacttatttttacataaaaaacgttttcattcattatttccatcattcattta
tctttctgtgttttttagtttagtttttagctagtttttttctaaattcctaacttta
aaaaatctggaaaagaaaattaaaaaattttgtccctattattttattttattact
ggaaaatcttcaaacaggaaaaccacgcgttttgcttattgctgtattttatga
aaaaaaaaacaatcaatattggtcaagtaaataagaaaaaattaacgaatctcta
tctgacaccagatgcgacctctattccacttctctgttcatctgctgcttcttt
tgtttaaccagataaatctccctcggggaaaaccgtcaaaaaaaggcaaaactaaa
tgcaaacacgctctatagacaaaatgtgtttggtctcgtcacgaatggtgagaga
gaattggcctccgccgcagagatcgcttgattattggcctccagtgggcaatgtc
ggggaaaaccaaactattgatgagagggtatcgacgaaaaatcaacaatgaccaac
tttttggttacagttttgttataaaatagagttttggatattccattgcgtatttt
tcttttctactttcaaaaaatctgctccaacctttaatggcttttccctgtcttgt
caaaatctggatttttgaatatataattttttaaaccatcaaattcagcgaaatg
aaatcatgtaatacaatttttttattttttccgactgttggtgtattccatcaaact
attcaaaaaatcaatataatgatttttttttcattttttcgcga

FIG. 5.

gagctcatgCGTcttcttgattactgtagatgtttggcaacttatttttacataa
aaacgttttcattcattatttccatcattcatttatctttctgtgtttttagtta
gttttagctagttttttctaaattcctaactttaaaaaatctggaaaagaaaat
taaaaaattttgtccctattattttattttattactggaaaatcttcaaacaggaa
aaccacCGcggttttgcttattgctgtatttatgaaaaaaaaaacaatcaatatt
ggtcaagtaaataagaaaaaattaacgaatctctatctgacaccagatgCGaccc
tctattccacttctctgttcatctgctgcttcttttggtttaaccagataaatctc
cctCGgggaaaaccgtcaaaaaaaggcaactaaatgcaaacacgctctatagac
aaaatgtgtttggtctCGtcacgaatggtgagagagaattggcctCGccCGcaga
gatCGcttgattattggcctccagtgggcaatgtCGgggaaaaccaaactattga
tgagaggatCGacgaaaaatcaacaatgaccaactttttgttacagttttgtta
taaatatgagttttggatattccattgCGtatttttcttttctactttcaaaaaa
tctgctccaacctttaatggcttttctgtcttgtcaaaatctggatttttgaat
atataatttttaaaacCATcaaattcagCGaaatgaaatcatgtaatacaatttt
ttattttttccgactgttgtgtattccatcaaactattcaaaaaatcaatataat
gatttttttttcatttttCGcgattttttattattttgtCGtctgaaaacctttt
tactaataaaaataatttacagggaaaaccactaacgactgtagccatgggaatca
gCGacaacgacgttcagaagcagctCGccCATgatggctttcattgagcaaga
ggccaatgagaaggctgaggagatCGatgctaaagCCgaggaagaattcaacatt
gagaaaagtaaggaattaaaacatttactcctttaaaactatactaaaaatctctc
taaaaaacggaaaaccttgaaattatgaattcattcaaattgtttcagggacgtc
ttgttcaacaacaacgtcaaaagattatggaattcttcgagaagaaggagaaaca
agtcgagcttcaacgcaaaattcaagcctccaactctctcaacgctggacgtctt
CGttgcttgaaaggTGagagaaaacgtttctcaacattttcaaaaacattaatCGc
cttaaaattgaaaaccagttctgaatCGgacacatttgaattaaaaacatatattt
caggctCGtgaaagaccacatCGgagCGtactCGacgaggctCGctCGaatctct
ccCGtattttCGgagatgctgctCGttatccagctattttgaagggacttgtcat
gcaaggacttcttcaattgctCGaaaaggaagtcgtccttCGttgCGgtgagaag
gatcttCGtcttgttgagcaacttttgccagagtgccttgacggacttcaaaagg
agtggggaagcaccaccaaggTCgttctCGataaacaacaaacttcttGCCatCGga
gtctgctgaggaggttgaactttctgctCGtgctggaagatccccgggattggc
ca

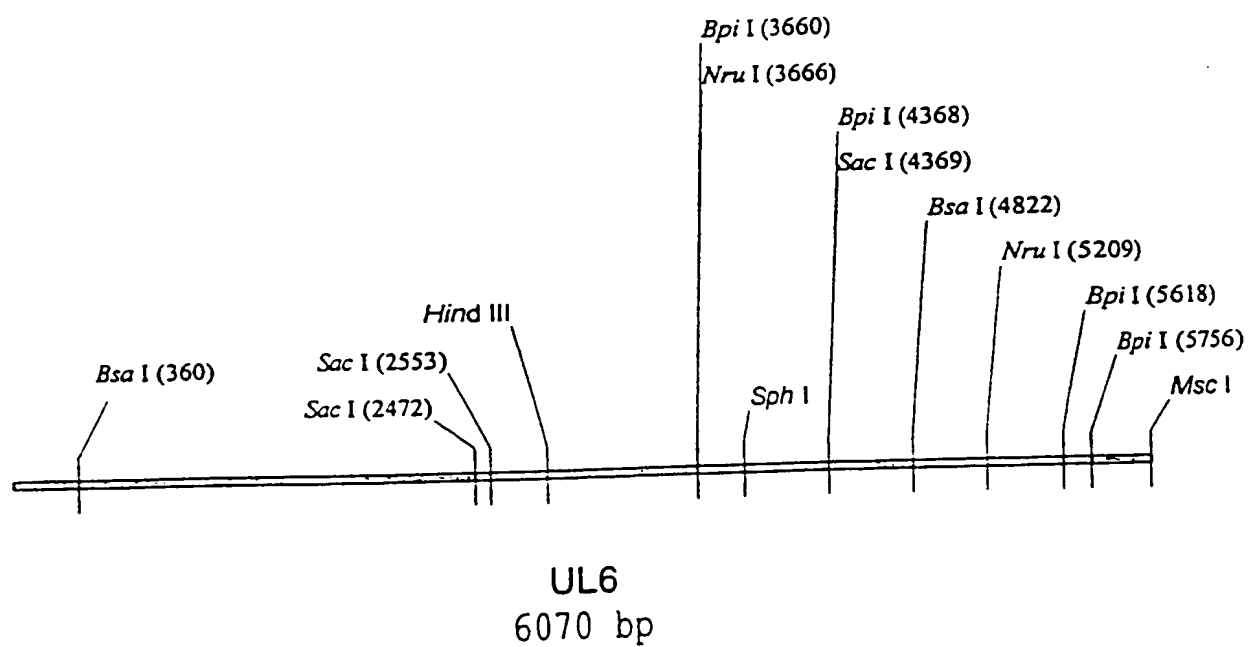
FIG. 6.

ttttcattcattatttccatcattcatttatctttctgtgttttttagtttagtttt
agctagtttttttctaaattcctaactttaaaaaatctggaaaagaaaattaaaa
aattttgtccctattatttattttattactggaaaatcttcaaacaggaaaaccc
accgcgttttgcttattgctgtatttatgaaaaaaaaaacaatcaatattggtca
agtaaataagaaaaaattaacgaatctctatctgacaccagatgcgaccctctat
tccacttctctgttcatctgctgcttcttttgtttaaccagataaaatctccctcg
gggaaaaccgtcaaaaaaaggcaaactaaatgcaaacacgctctatagacaaaat
gtgtttggtctcgtcacgaatggtgagagagaattggcctccgccgcagagatcg
cttgattattggcctccagtgggcaatgtcggggaaaaccaaactattgatgaga
ggtatcgacgaaaaatcaacaatgaccaactttttgttacagttttgttataaat
atgagttttggatattccattgcgtatttttcttttctactttcaaaaaatctgc
tccaacctttaatggcttttctgtcttgtcaaaatctggatttttgaatatata
atttttaaaaccatcaaattcagcgaaatgaaatcatgtaatacaattttttatt
ttttccgactgttgtgtattccatcaaactattcaaaaaatcaatataatgattt
ttttttcatttttcgcgat

FIG. 7.

ggtctcgtcacgaatggtgagagagaattggcctccgccgcagagatcgcttgat
tattggcctccagtgggcaatgtcggggaaaaccaaactattgatgagagggtatc
gacgaaaaatcaacaatgaccaactttttgttacagttttgttataaatatgagt
tttgatattccattgcgtatttttcttttctactttcaaaaaatctgctccaac
ctttaatggcttttctgtcttgtcaaaatctggatttttgaatatataattttt
aaaaccatcaaattcagcgaaatgaaatcatgtaatacaattttttattttttcc
gactgttgtgtattccatcaaactattcaaaaaatcaatataatgatttttttt
catttttcgcga

FIG. 8.



GFP

+	
+	
+	
-	
+	
+	
-	
+	
+	
-	

UL6 (6 kb)	pGF2002 (3.2 kb)	pGF2003 (2.2 kb)	pGF2001 (0.9 kb)	pGF2006 (1.3 kb)	pGP2008 (1.7 kb)	pGF2010 (0.5 kb)	pGF2009 (0.8 kb)	pGF2012 (0.4 kb)	pGF2011 (0.5 kb)
------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------	------------------

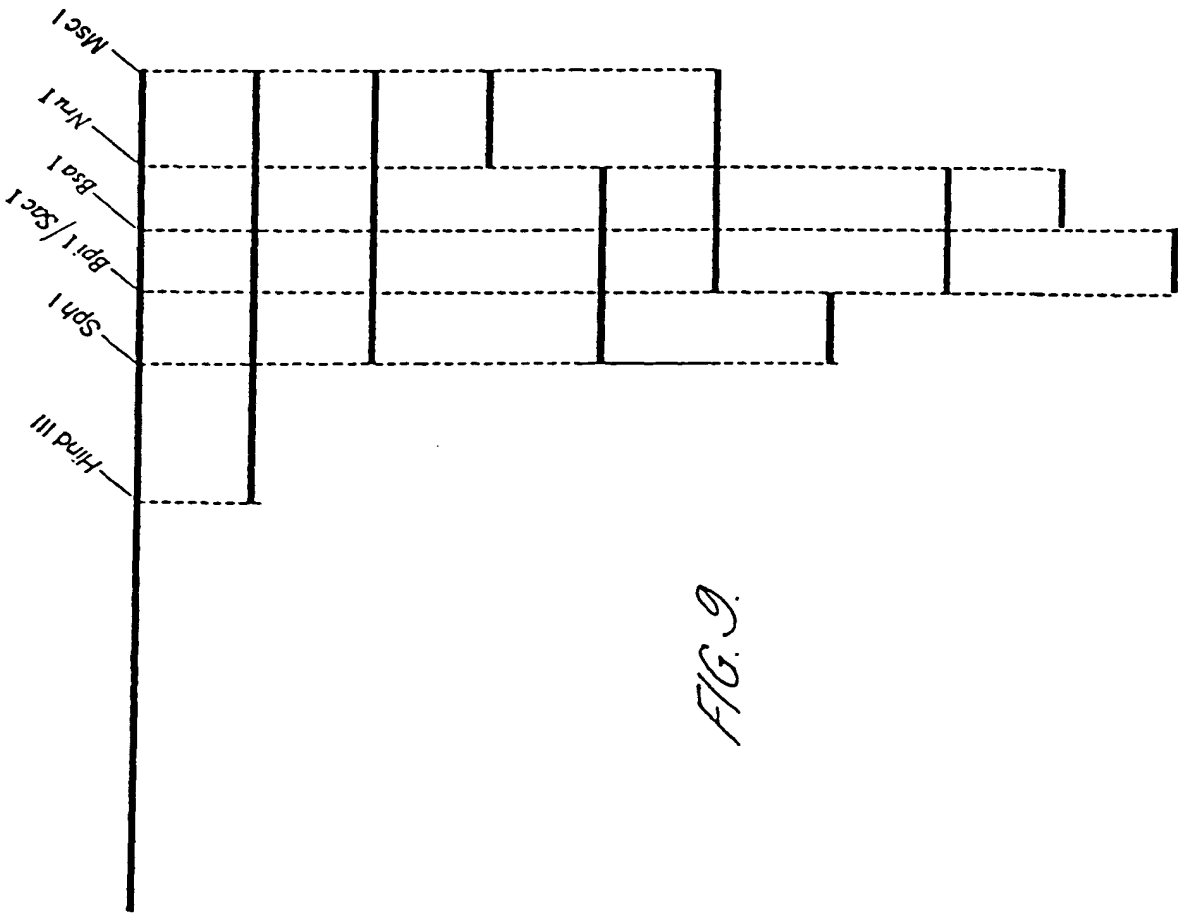


FIG. 9.

FIG. 10.

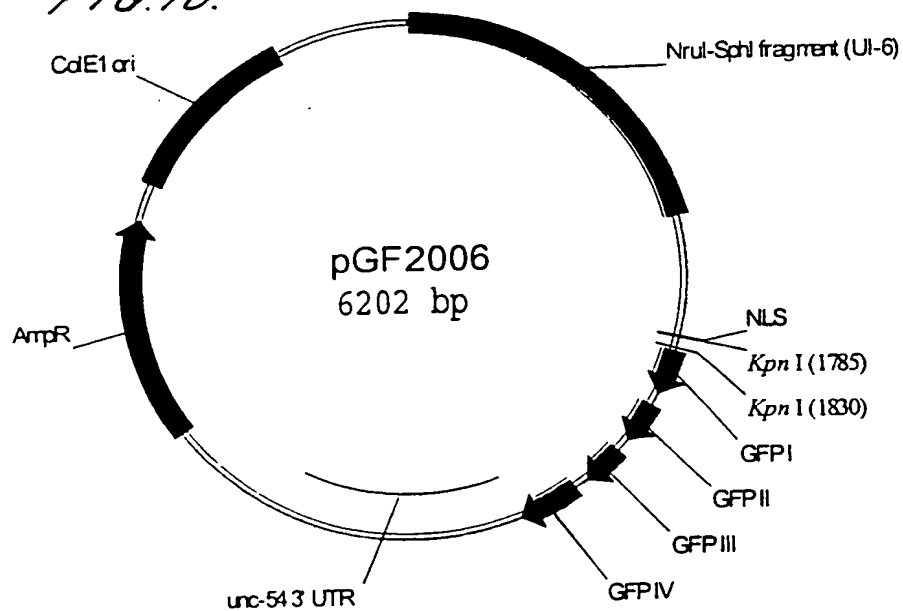


FIG. 11.

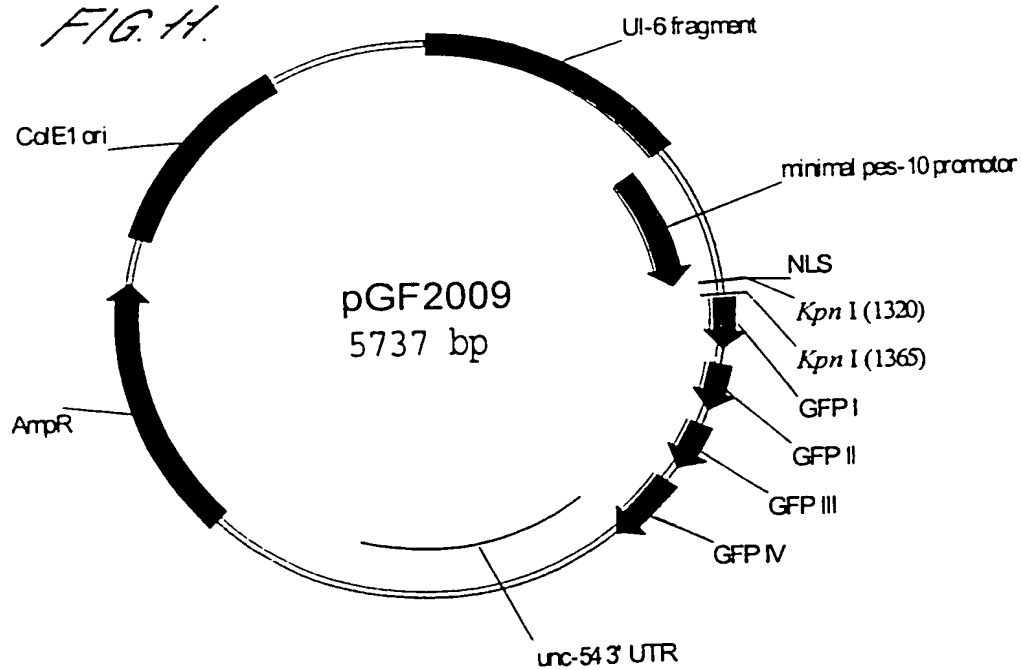


FIG. 12.

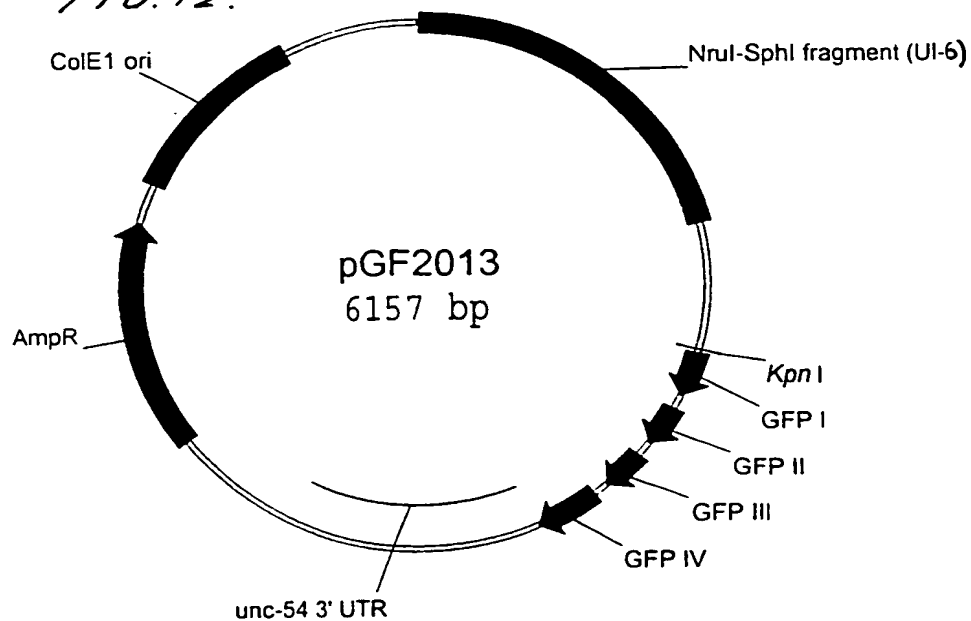
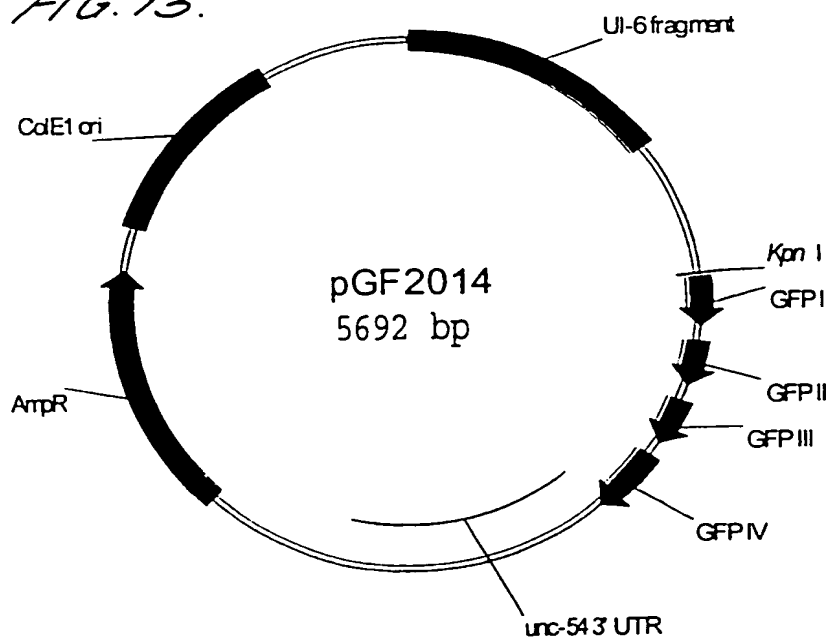


FIG. 13.



SEQUENCE LISTING

<110> DE'GEN NV

<120> Expression of DNA or proteins in *C. elegans*

<130> SCB/51936/002

<140>

<141>

<160> 11

<170> PatentIn Ver. 2.0

<210> 1

<211> 6070

<212> DNA

<213> *Caenorhabditis elegans*

<400> 1

```
gatcgctcga ataaaaaatt ttataatgtc agattacgtt ttagatccaa aaaaaattag 60
gctgaatccg tttttgattc tctagaattt tgttttagcca aaaaacatcg cattgttctc 120
tctgtcgtta ttaaagtgcg caattaacaa atattcaaaa attcttattt aaaaaacat 180
aggggggttc ctaactattg cgaaattctc tttctctctc ggacattact atgagagatt 240
gaaccaattg aagagacgca gtgctaggaa gggaaatgcg cctggcgcca catatgggtc 300
aagtctctct gaatcgtgtg ctctctacat tgggtcgact ctgcgagtgc ttttgagcgg 360
cgccgagacc gaaataagag agaaaaacaga gggaaaaaag atagggaaca ttaattgatg 420
aaaaaagaaa cagatgacgt ggcaattctt ctttccggag ttttttttta ttggaagtgg 480
ggttattcaa gtaatgtagc aaaatgtatt taaatacatt tgtgacgtca caaatgtatt 540
taaatacatg tttttatata cttgaataag gttgtgacgt aatttttcta cactttttaa 600
ttttccgaca ctacttgaat aaccccaaaa gtgtacggtt tcttttttca aaacacgatt 660
gcaaccaaag gcgcgggatt tttgaatttt tcaaaaatcg gaagatttaa aattttcgct 720
tttttatctt tattacttga taaaaattga attttatggt gaaatttcaa aatttgtttt 780
ggctttttcag ttgtccggaa aatcgaaaaa gtttgttttt cgattccac tgaaaaaatc 840
gaattttttc gatgaaacct ttgattgcaa tgttagaaaa atttctaaaa aattaacaag 900
aacactggaa caacacaaaa tcacaataat caacaccgga ccattgggtac cgtataaaaa 960
gaaaaaaaaa tcgaattaat tgtaatatgt ctgaaaaatc ttcataaata cccattttga 1020
atataattat ttcttaataa atatataact tacagagaaa aaaatttcgg ggacgttaag 1080
gctcatagaa gaacacacga aaattaggta agttaaggct gaaggcattt ggcctactct 1140
acacgtagca gggaaatgat gtaaaatgga aaagagagaa aatatataaa acatatgaat 1200
gtgcaaggat tcggtatata aattatataa gaagggtcct tattaaggaa tttgttttcc 1260
tgtatgtaca atctagaaaa gagcagaagg gttattgaaa ggggtggagta gagtcaaaat 1320
tgttttttta ctttaaatga cagaaaatgg acactaatga cggaatataa ctataaacat 1380
tttttctaaa ttttgaaaaa tgattttttc aattttgcac ctctcaaatt aagagcacta 1440
aagaaaatta cagaaaaacc agtaaaattt gagaatttga aacggatatc tccaaaaagt 1500
agtcgaagtc taagaatcaa aaaagaaatt aaaatttttt ttagttacgg tatattttct 1560
gtcatttaaa gcagttgact ccactccacc tttaaaacaa tataaattat ctaatgaggc 1620
ctagcagaat atcggttcct caaatacgga ttatagggtg tcatatcgcc ttttctccat 1680
tcgctccaaa cgatttttggc gacaaccaga tgaaaagcga cgacgacaac aaaaacgaac 1740
accagaatg ttgtcatttc tccagttgat tctccagatc gatgcatttc ataacttct 1800
tcttcgtgtg ccggtgctgc accaacttcc gacacgtgaa tggcatttct taggatttca 1860
agatgtcgtc gagatggagc aatgtagaga agagatgtta ctgggttctg gacttttact 1920
ggtaattgaa ctggagtctt tccaagtga aactgggctat cgaggcatcc accatcacia 1980
acaaataatt gactagaaga tgaaacctga aagaaatagt ttagagtatt tccaggataa 2040
atatttttaa aatttaaaata aaagacgtgt ttcagctctt ttgacaagga tttttattca 2100
ttttttgaga agtctcacta tgaaaattcg ggcagtttca gtctatttaa gcaacaagaa 2160
accacaaact atgtacagat tgtaaaact tctataaact attaccataa aataaggacg 2220
attttccgta tccaatttta tatccactcg aatagtagca acaacacca tattcgaata 2280
aactggaatt ccatgaaaaa tcatttcacg atgcaaatca tgagcatcca atgtacttc 2340
```

caaatgagaa	tcgtgataac	tcgatgcagt	caacgattga	acaaatgctt	gtgatgctcc	2400
aattgctcca	gtgctcaaaa	gattgttgca	acctcgtttt	tggaaaagtga	gcattccatgt	2460
attcacgagc	tcaaaaagtt	ttgtcggagt	atcccagtcct	tgccacagac	gtgatgggaa	2520
gagaagattt	gaatgaccag	tgtagcagag	ctgcgttttc	cgcgccaaat	ttccagagg	2580
ctataaacat	tctattttaa	ttaggtaata	atattcaatt	acctttacag	attcnaaagt	2640
agccccatgc	tcaattgttg	gagcccgtaa	attgctcatt	aaaatgtatt	ttgttgcat	2700
aattctatct	ccgtttaatg	tatttggtgc	tttcgaattc	gcaacactga	atcctgttaa	2760
atgcatcttg	ctccatccgt	caaaaatgctc	cgaatccaaa	gttttcgagt	cttttcctga	2820
ttggaaaaat	cggatagcat	cccgagaagc	ttcttcaact	gttggcttcg	acgcaatcga	2880
gcattgtaaac	cgaaatgact	cttctctgaa	ataaactgaa	caaatttaatt	aatttttttt	2940
atgttttaaa	tatacctttt	ctgtgaaact	gaaacttttc	cagggtgcac	agtacaaaaca	3000
acagcatatc	atgctctagc	aacttgccgc	tgatgaaccg	tcagctcgcc	taattttgaa	3060
tgagccacc	agttggaact	aagacgagac	aagttgagtc	ggaccgttga	tttctgaaat	3120
taaatgttct	ttttgtaatt	ttaatgcac	aatttcaata	aaattacagt	tggaattggaa	3180
acttgtatca	attgaacaaa	atagtttgga	cgagttcgat	gagcccatat	ctgtgacgtc	3240
acacatgcac	attcgccatc	ctgaaacatt	aatgtccaat	aaatagttta	ttatgtcgtg	3300
gctaaccatt	gaaaaacatc	gcagaatttt	acaaactcca	tttttgaaat	ccgtcatttt	3360
tgtaacctca	acaggttggt	gtccgtcaat	ttcaatgttg	agcccgggat	aaaaattggg	3420
ttccacatcg	agaactcggt	ttttctcctg	aactcgtaat	tgccgaatg	aatcaacacc	3480
aatatatcca	tttccacaaa	actctacgct	ttgattgtta	attcctgtat	cgaactccta	3540
aaaaaaaagta	aatatgaata	attactaatt	attcgactca	catcaatctc	tttttcccat	3600
aatttttaact	tttccgctga	acattgatgc	tgttttgatga	atgatgtgcc	ggaagacgtt	3660
cctcgcgaga	agatgaaaat	gacgaataaa	acaaactcca	taattatcgc	taatcttcgt	3720
tttgatatc	gtcgaaccat	acgcatttta	ctttcgaaat	ccatattttc	tattcgacgc	3780
acgaaagtgc	gccacgttat	tgcggcgctc	attttgagg	ggaaatccgc	ctgaaaaatc	3840
aattgtttgt	attgtgaaat	ttcgaagagg	cataaaacaa	gaaaacggac	atgaaagcgc	3900
gttgcatgca	aggttagttg	cctgtttaag	cattatcccc	gcattgtagct	tgttcggcac	3960
cgttaaaaat	gctgagtaat	cagcttttta	gaatttaaaa	tattaaactt	ttaaaattgc	4020
aacaaacatc	gacaaatatt	caagaggcga	atgatatcgg	gaatttcgat	tgaaacgaaa	4080
ctgttttgaa	attcaaaaag	tattttcaaa	gtattgtccg	caaggcacat	cacgcaaact	4140
tgcagaatct	accgtatccc	atacattttt	atagtttttc	cctcagtttt	taaaattaaa	4200
aacgctgaaa	aagcgattaa	atttatttta	atgcatcgtt	cgaataaaaat	aaagtttatc	4260
ttttgatata	aacatgagtt	tcccttgagg	aaaagtaggg	atttcgcctt	tcaaaaaatt	4320
atttcgtgca	ggatgctatt	ttcgtggcga	aaccatact	caagagctca	tgcgctctct	4380
tgattactgt	agatgtttgg	caacttattt	ttacataaaa	acgttttcat	tcattatttc	4440
catcttctat	ttatctttct	gtgtttttag	ttagttttag	ctagtttttt	tctaaattcc	4500
taacttttaa	aaatctggaa	aagaaaaatta	aaaaattttg	tccctattat	ttattttatt	4560
actgaaaaat	cttcaaacag	gaaaacccac	cgcgttttgc	ttattgctgt	atztatgaaa	4620
aaaaaaaacaa	tcaatattgg	tcaagtaaat	aagaaaaaat	taacgaatct	ctatctgaca	4680
ccagatgcga	ccctctatct	cacttctctg	ttcatctgct	gcttcttttg	tttaaccaga	4740
taaattctcc	tcggggaaaa	ccgtcaaaaa	aaggcaaaact	aaatgcaaac	acgctctata	4800
gacaaaaatgt	gtttggtctc	gtcacgaatg	gtgagagaga	attggcctcc	gccgcagaga	4860
tcgcttgatt	attggcctcc	agtgggcaat	gtcggggaaa	accaaactat	tgatgagagg	4920
tatcgacgaa	aaatcaacaa	tgaccaactt	tttgttacag	ttttgttata	aatatgagtt	4980
ttggatatct	cattgcgtat	ttttcttttc	tactttcaaa	aaatctgctc	caacctttta	5040
tggtcttttc	tgtcttgta	aaatctggat	ttttgaatat	ataattttta	aaacctatca	5100
attcagcgaa	atgaaatcat	gtaatacaat	tttttatttt	ttccgactgt	tgtgtatttc	5160
atcaaaactat	tcaaaaaatc	aatataatga	tttttttttc	attttttcg	atttttttat	5220
attttgctgt	ctgaaaaact	ttttactaat	aaaataattt	acagggaaaa	ccactaacga	5280
ctgtagccat	gggaatcagc	gacaacgacg	ttcagaagca	gctccgccac	atgatggctt	5340
tcattgagca	agaggccaat	gagaaggctg	aggagatcga	tgctaaagcc	gaggaagaat	5400
tcaacattga	gaaagttaag	aattaaaaaca	tttactcctt	taaaactata	ctaaaaatctc	5460
ttctaaaaaa	cggaaaaact	tgaaattatg	aattcattca	aattgtttca	gggacgtctt	5520
gttcaaaaac	acgttcaaaa	gatttatggaa	ttcttcgaga	agaaggagaa	acaagtcgag	5580
cttcaacgca	aaattcaagc	ctccaactct	ctcaacgctg	gacgtcttcg	ttgcttgaag	5640
gtgagagaaa	acgtttctca	acattttcaa	aaacattaat	cgccttaaaa	ttgaaaacca	5700
gttctgaatc	ggacacattt	gaattaaaaa	catattttca	ggctcgtgaa	gaccacatcg	5760
gagccgtact	cgacgaggtt	cgctcgaatc	tctcccgat	ttccggagat	gctgctcgtt	5820
atccagctat	tttgaaggga	cttgtcatgc	aaggacttct	tcaattgctc	gaaaagggaag	5880
tcgtccttcg	ttgccgtgag	aaggatcttc	gtcttggtga	gcaacttttg	ccagagtgc	5940

```

ttgacggact tcaaaaggag tggggaagca ccaccaaggt cgttctcgat aaacaaaact 6000
tcttgccatc ggagtcctgct ggaggagttg aactttctgc tcgtgctgga aagatccccg 6060
ggattggcca                                     6070

```

<210> 2

<211> 3224

<212> DNA

<213> *Caenorhabditis elegans*

<400> 2

```

aagcttcttc acttgttggc ttcgacgcaa tcgagcatgt aaaccgaaat gactcttctc 60
tgaaataaac tgaacaaaatt aattaatttt tttaatgttt taaatatacc ttttctgtga 120
aactgaaact tttccagggt catcagtaga aacaacagca tacgatgctc tagcaacttg 180
ccgctgatga accgtcagct cgcctaattt tgaatgagcc caccagttgg aactaagacg 240
agacaagtgg agtcggaccg ttgatttctg aaattaaatg ttctttttgt aatttttaag 300
catcaatttc aataaaaatta cagttggatt ggaaacttgt atcaattgaa caaaatagtt 360
tggacgagtt cgatgagccc atatctgtga cgtcacacat gcacattcgc catcctgaaa 420
cattaatgtc caataaatag tttattatgt cgtggctaac cattgaaaaa catcgacgaa 480
tttgaaaaat gccatttttg aaatccgtca tttttgtaac ctcaacaggt tgttgtccgt 540
caatttcaat gttgagcccc ggataaaaaa tggtttccac atcgagaact cggtttttct 600
cctgaactcg taattgtccg aatgaatcaa caccaatata tccatttccc acaaactcta 660
cgctttgatt gttaattcct gtatcgaaact cctaaaaaaa agtaaatatg aataattact 720
aattattcga ctcacatcaa tctcttttcc ccataatttt aacttttccg ctgaacattg 780
atgtctgttt gtgaatgatg tgccggaaaga cgttcctcgc gagaagatga aaatgacgaa 840
taaaacaact ccaataatta tcgctaattc tcgttttggg tatcgtcgaa ccatacgcat 900
tttactttcg aaatccatat tttctattcg acgcacgaaa gtgcgccacg ttattgcggc 960
gctcattttg gaggggaaat ccgcctgaaa aatcaattgt ttgtattgtg aaatttcgaa 1020
gaggcataaa acaagaaaac ggacatgaaa gcgcgttgca tgcaaggtta gttgcctgtt 1080
taagcattat ccccgcatgt agcttgctcg gcaccgttaa aaatgctgag taatcagctt 1140
tttagaattt aaaatattaa acttttaaaa ttgcaacaaa catcgacaaa tattcaagag 1200
gcgaatgata tcgggaattt cgattgaaac gaaactgttt tgaaattcaa aaagtatttt 1260
caaagtattg tccgcaaggc acatcacgca aacttgcaag atctaccgta tcccatacat 1320
ttttatagtt tttccctcag tttttaaaaa taaaaacgct gaaaaagcga ttaaatttat 1380
ttaaatgcat cgttcgaaata aaataaagtt tatcttttga taaaaacatg agtttccttt 1440
ggagaaaagt agggatttcg cctttcaaaa aattatttcg tgcaggatgc tattttcgtg 1500
gcgaaaccca tactcaagag ctcatgcgtc tcttgattta ctgtagatgt ttggcaactt 1560
atttttacat aaaaacgttt tcattcatta tttccatcat tcatttatct ttctgtgttt 1620
ttagttagtt ttagctagtt tttttctaaa ttcttaactt taaaaaatct ggaaaagaaa 1680
attaaaaaat tttgtcccta ttattcattt tattactgga aaatcttcaa acaggaaaaa 1740
ccaccgcgtt ttgcttattg ctgtatttat gaaaaaaaaa acaatcaata ttggtcaagt 1800
aaataagaaa aaattaacga atctctatct gcaccagat gcgacctctc attccactct 1860
tctgttcacg tgctgttctc tttgtttaac cagataaaatc tccctcgggg aaaaccgtca 1920
aaaaaaggca aactaaatgc aaacacgctc tatagacaaa atgtgtttgg tctcgtcacg 1980
aatggtgaga gagaattggc ctccgcccga gagatcgctt gattattggc ctccagtggg 2040
caatgtcggg gaaaacccaa ctattgatga gaggtatcga cgaaaaatca acaatgacca 2100
actttttgtt acagttttgt tataaatatg agttttggat attccattgc gtatttttct 2160
tttctacttt caaaaaatct gctccaacct tttaattgctt ttctgtctt gtcaaaaatct 2220
ggatttttga atatataatt tttaaaaacca tcaaattcag cgaaatgaaa tcatgtaata 2280
caatttttta ttttttccga ctgttgtgta ttccatcaaa ctattcaaaa aatcaatata 2340
atgatttttt tttcattttt cgcgattttt tattattttg tcgtctgaaa acctttttac 2400
taataaaaata atttacaggg aaaaccacta acgactgtag ccatgggaat cagcgacaac 2460
gacgttcaga agcagctccg ccacatgatg gctttcattg agcaagaggc caatgagaag 2520
gctgaggaga tctgatgctaa agccgaggaa gaattcaaca ttgagaaagt aaggaattaa 2580
aacatttact cctttaaacc tatactaaaa tctcttctaa aaaacggaaa accttgaaat 2640
tatgaattca ttcaaattgt ttcagggacg tcttgttcaa caacaacgctc aaaagattat 2700
ggaattcttc gagaagaagg agaaacaagt cgagcttcaa cgcaaaattc aagcctccaa 2760
ctctctcaac gctggacgtc ttcgttgctt gaagtgaga gaaaacgttt ctcaacattt 2820
tcaaaaacat taatcgctt aaaaattgaaa accagttctg aatcggacac atttgaaata 2880
aaaaacatat ttcaggctcg tgaagaccac atcggagccg tactcgacga ggctcgctcg 2940
aatctctccc gtatttccgg agatgctgct cgttatccag ctattttgaa gggacttgtc 3000

```

```

atgcaaggac ttcttcaatt gctcgaaaag gaagtcgtcc ttcgttgcgc tgagaaggat 3060
cttcgtcttg ttgagcaact tttgccagag tgccttgacg gacttcaaaa ggagtgggga 3120
agcaccacca aggtcgttct cgataaacia aacttcttgc catcggagtc tgctggagga 3180
gttgaaacttt ctgctcgtgc tggaaaagatc cccgggattg gccca 3224

```

<210> 3

<211> 2167

<212> DNA

<213> *Caenorhabditis elegans*

<400> 3

```

gcatgcaagg ttagttgcct gtttaagcat tatccccgca ttagcttgt tccgcaccgt 60
taaaaatgct gagtaatcag ctttttagaa tttaaaatat taaactttta aaattgcaac 120
aaacatcgac aaatattcaa gaggcgaatg atatcgggaa tttcgattga aacgaaactg 180
ttttgaaatt caaaaagtat tttcaaagta ttgtccgcaa ggcacatcac gcaaacttgc 240
agaatctacc gtatcccata cttttttata gtttttccct cagttttttaa aattaaaaaac 300
gctgaaaaag cgattaaatt tatttaaatt catcgttcga ataaaaataa gtttatcttt 360
tgataaaaaac atgagtttcc tttggagaaa agtagggatt tcgcctttca aaaaattatt 420
tcgtgcagga tgctattttc gtggcgaaac ccatactcaa gagctcatgc gtcttcttga 480
ttactgtaga tgtttgcaa cttattttta cataaaaaacg ttttcattca ttatttccat 540
cattcattta tcttctgtg ttttttagtta gtttttagcta gtttttttct aaattcctaa 600
ctttaaaaaa tctggaaaag aaaattaaaa aattttgtcc ctattattta ttttattact 660
ggaaaatctt caaacaggaa aaccacccgc gttttgctta ttgctgtatt tatgaaaaaa 720
aaaacaatca atattggtca agtaataaag aaaaaattaa cgaatctcta tctgacacca 780
gatgcgaccc tctattccac ttctctgttc atctgtgct tcttttgtt aaccagataa 840
atctccctcg gggaaaaccg tcaaaaaaag gcaaaactaaa tgcaaacacg ctctatagac 900
aaaatgtgtt tggctctcgtc acgaatggtg agagagaatt ggcctccgcc gcagagatcg 960
cttgattatt ggcctccagt gggcaatgtc ggggaaaacc aaactattga tgagaggtat 1020
cgacgaaaaa tcaacaatga ccaacttttt gttacagttt tggtataaat atgagttttg 1080
gatattccat tgcgtatttt tcttttctac tttcaaaaaa tctgctccaa cttttaatgg 1140
cttttctgtt ctgttcaaaa tctggatttt tgaatatata atttttaaaa ccatcaaatt 1200
cagcgaaatg aaatcatgta atacaatttt ttatttttct cgactgttgt gtattccatc 1260
aaactattca aaaaatcaat ataattgattt ttttttcatt tttcgcgatt ttttattatt 1320
ttgtcgtctg aaaacctttt tactaataaaa ataatttaca gggaaaacca ctaacgactg 1380
tagccatggg aatcagcgac aacgacgttc agaagcagct ccgccacatg atggctttca 1440
ttgagcaaga ggccaatgag aaggctgagg agatcgatgc taaagccgag gaagaattca 1500
acattgagaa agtaaggaat taaaacattt actcctttta aactatacta aaatctcttc 1560
taaaaaacgg aaaaccttga aattatgaat tcatcctaaa tgtttcaggg acgtcttgtt 1620
caacaacaac gtcaaaaagat tatggaattc ttcgagaaga aggagaaaca agtcgagctt 1680
caacgcaaaa ttcaagcctc caactctctc aacgctggac gtcttcgttg cttgaagggtg 1740
agagaaaaacg tttctcaaca ttttcaaaaa cattaatcgc cttaaaattg aaaaccagtt 1800
ctgaatcgga ccaatttgaa ttaaaaacat attttcaggc tcgtgaagac cacatcggag 1860
ccgtactcga cgaggctcgc tcgaatctct cccgtatttc cggagatgct gctcgttatc 1920
cagctatttt gaagggactt gtcattgcaag gacttcttca attgctcgaa aaggaagtcg 1980
tccttcgttg ccgtgagaag gatcttcgtc ttgttgagca acttttgcca gagtgccttg 2040
acggacttca aaaggagtgg ggaagcacca ccaaggctcg tctcgataaa caaaacttct 2100
tgccatcgga gtctgctgga ggagttgaac tttctgctcg tgctggaaaag atccccggga 2160
ttggcca 2167

```

<210> 4

<211> 1308

<212> DNA

<213> *Caenorhabditis elegans*

<400> 4

```

gcatgcaagg ttagttgcct gtttaagcat tatccccgca ttagcttgt tccgcaccgt 60
taaaaatgct gagtaatcag ctttttagaa tttaaaatat taaactttta aaattgcaac 120
aaacatcgac aaatattcaa gaggcgaatg atatcgggaa tttcgattga aacgaaactg 180
ttttgaaatt caaaaagtat tttcaaagta ttgtccgcaa ggcacatcac gcaaacttgc 240
agaatctacc gtatcccata cttttttata gtttttccct cagttttttaa aattaaaaaac 300

```

```

gctgaaaaag cgattaaatt tatttaaatt catcggttca ataaaaataaa gtttatcttt 360
tgataaaaac atgagtttcc ttgggagaaa agtagggatt tcgcctttca aaaaattatt 420
tcgtgcagga tgctattttc gtggcgaaac ccatactcaa gagctcatgc gtcttcttga 480
ttactgtaga tgtttgga caatttttta cataaaaacg ttttcattca ttatttccat 540
cattcattta tctttctgtg tttttagtta gtttttagcta gttttttttt aaatttctaa 600
ctttaaaaaa tctggaaaag aaaattaaaa aattttgtcc ctattattta ttttattact 660
ggaaaatctt caaacaggaa aaccaccgc gttttgctta ttgctgtatt tatgaaaaaa 720
aaaaaatca atattggtca agtaataaag aaaaaattaa cgaatctcta tctgacacca 780
gatgcgaccc tctattccac ttctctgttc atctgctgct tcttttgttt aaccagataa 840
atctccctcg gggaaaaccg tcaaaaaaag gcaaaactaa tgcaaacacg ctctatagac 900
aaaatgtgtt tggctcgtc acgaatggtg agagagaatt ggcctccgc gcagagatcg 960
cttgattatt ggcctccagt gggcaatgtc ggggaaaacc aaactattga tgagaggtat 1020
cgacgaaaaa tcaacaatga ccaacttttt gttacagttt tggtataaat atgagttttg 1080
gatattccat tgcgtatttt tcttttctac tttcaaaaaa tctgctccaa cctttaatgg 1140
cttttctgt cttgtcaaaa tctggatttt tgaatatata atttttaaaa ccatcaaatt 1200
cagcgaaatg aaatcatgta atacaatttt ttattttttc cgactgttgt gtattccatc 1260
aaactattca aaaaatcaat ataatgattt ttttttcatt tttcgca 1308

```

<210> 5

<211> 1707

<212> DNA

<213> *Caenorhabditis elegans*

<400> 5

```

gagctcatgc gtcttcttga ttactgtaga tgtttgga caatttttta cataaaaacg 60
ttttcattca ttatttccat cattcattta tctttctgtg ttttttagtta gtttttagcta 120
gtttttttct aaatttctaa ctttaaaaaa tctggaaaag aaaattaaaa aattttgtcc 180
ctattattta ttttattact ggaaaatctt caaacaggaa aaccaccgc gttttgctta 240
ttgctgtatt tatgaaaaaa aaaacaatca atattggtca agtaataaag aaaaaattaa 300
cgaatctcta tctgacacca gatgcgaccc tctattccac ttctctgttc atctgctgct 360
tcttttgttt aaccagataa atctccctcg gggaaaaccg tcaaaaaaag gcaaaactaa 420
tgcaaacacg ctctatagac aaaatgtgtt tggctcgtc acgaatggtg agagagaatt 480
ggcctccgc gcagagatcg cttgattatt ggcctccagt gggcaatgtc ggggaaaacc 540
aaactattga tgagaggtat cgacgaaaaa tcaacaatga ccaacttttt gttacagttt 600
tggtataaat atgagttttg gatattccat tgcgtatttt tcttttctac tttcaaaaaa 660
tctgctccaa cctttaatgg cttttctgtt cttgtcaaaa tctggatttt tgaatatata 720
atttttaaaa ccatcaaatt cagcgaaatg aaatcatgta atacaatttt ttattttttc 780
cgactgttgt gtattccatc aaactattca aaaaatcaat ataatgattt ttttttcatt 840
tttcgcgatt ttttattatt ttgtcgtctg aaaacctttt tactaataaa ataatttaca 900
gggaaaacca ctaacgactg tagccatggg aatcagcgac aacgacgttc agaagcagct 960
ccgccacatg atggctttca ttgagcaaga ggccaatgag aaggctgag agatcgatgc 1020
taaaagccgag gaagaattca acattgagaa agtaaggaa taaaacattt actcctttaa 1080
aactatacta aaatctcttc taaaaaacgg aaaaccttga aattatgaat tcattcaaatt 1140
tgtttcaggg acgtcttgtt caacaacaac gtcaaaaagat tatggaattc ttcgagaaga 1200
aggagaaaca agtcgagctt caacgcaaaa ttcaagcctc caactctctc aacgctggac 1260
gtcttcgttg cttgaagggt agagaaaaacg ttctcaaca ttttcaaaaa cattaatcgc 1320
cttaaaaattg aaaaccagtt ctgaatcgga cacatttgaa ttaaaaacat attttcaggc 1380
tcgtgaagac cacatcgag ccgtactcga cgaggctcgc tcgaatctct cccgtatttc 1440
cgagatgct gctcgttatc cagctatttt gaagggactt gtcatgcaag gacttcttca 1500
attgctcgaa aaggaaatcg tcttctgtt ccgtgagaag gatcttcgtc ttgttgagca 1560
acttttgcca gagtgccttg acggacttca aaaggagtgg ggaagcacca ccaaggctcg 1620
tctcgataaa caaaacttct tgccatcgga gtctgctgga ggagttgaac tttctgctcg 1680
tgctggaaag atccccggga ttggcca 1707

```

<210> 6

<211> 789

<212> DNA

<213> *Caenorhabditis elegans*

<400> 6

```

ttttcattca ttatttccat cattcattta tctttctgtg tttttagtta gtttttagcta 60
gtttttttct aaattcctaa ctttaaaaaa tctggaaaag aaaattaaaa aattttgtcc 120
ctattattta ttttattact ggaaaatctt caaacaggaa aaccaccgc gttttgctta 180
ttgctgtatt tatgaaaaaa aaaacaatca atattggtca agtaaataag aaaaaattaa 240
cgaatctcta tctgacacca gatgcgaccc tctattccac ttctctgttc atctgctgct 300
tcttttgttc aaccagataa atctccctcg gggaaaaccg tcaaaaaaag gcaaaactaaa 360
tgcaaacacg ctctatagac aaaatgtgtt tggctcgtc acgaatggtg agagagaatt 420
ggcctccgcc gcagagatcg cttgattatt ggcctccagt gggcaatgtc ggggaaaacc 480
aaactattga tgagagggtat cgacgaaaaa tcaacaatga ccaacttttt gttacagttt 540
tggtataaat atgagttttg gatattccat tgcgtatttt tcttttctac tttcaaaaaa 600
tctgctccaa cctttaatgg cttttcctgt cttgtcaaaa tctggatttt tgaatatata 660
atttttaaaa ccatcaaat cagcgaaatg aaatcatgta atacaatttt ttattttttc 720
cgactgttgt gtattccatc aaactattca aaaaatcaat ataatgattt ttttttcatt 780
tttcgcgat                                     789

```

<210> 7

<211> 397

<212> DNA

<213> *Caenorhabditis elegans*

<400> 7

```

ggctctgtca cgaatggtga gagagaattg gcctccgccg cagagatcgc ttgattattg 60
gcctccagtg ggcaatgtcg gggaaaacca aactattgat gagaggatc gacgaaaaat 120
caacaatgac caactttttg ttacagtttt gttataaata tgagttttgg atattccatt 180
gcgtattttt cttttctact ttcaaaaaat ctgctccaac ctttaatggc ttttcctgtc 240
ttgtcaaaat ctggattttt gaatatataa tttttaaaac catcaaattc agcgaaatga 300
aatcatgtaa tacaattttt tattttttcc gactgttgtg tattccatca aactattcaa 360
aaaatcaata taatgatttt tttttcattt ttccgca                                     397

```

<210> 8

<211> 6202

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: plasmid
pGF2006

<400> 8

```

agcttgcatg caaggttagt tgcctgttta agcattatcc ccgcatgtag cttgttcggc 60
accgttaaaa atgctgagta atcagctttt tagaatttaa aatattaaac ttttaaaatt 120
gcaacaaaac atgacaaaata ttcaagaggc gaatgatatc gggaatttcg attgaaacga 180
aactgttttg aaattcaaaa agtattttca aagtattgtc cgcaaggcac atcacgcaaa 240
cttgacagaat ctaccgtatc ccatacattt ttatagtttt tccctcagtt tttaaaatta 300
aaaacgctga aaaagcgatt aaattttatt aaatgcatcg ttcgaataaa ataaagttta 360
tcttttgata aaaacatgag tttccttttg agaaaagtag ggatttcgcc tttcaaaaaa 420
ttatttcgtg caggatgcta ttttcgtggc gaaaccata ctcaagagct catgctgctt 480
cttgattact gtagatgttt ggcaacttat ttttacataa aaacgttttc attcattatt 540
tccatcattc atttatcttt ctgtgttttt agttagtttt agctagtttt tttctaaatt 600
cctaacttta aaaaatctgg aaaagaaaat taaaaaattt tgtccctatt atttatttta 660
ttactggaaa atcttcaaac aggaaaaccc accgcgtttt gcttattgct gtatttatga 720
aaaaaaaaac aatcaatatt ggtcaagtaa ataagaaaaa attaacgaat ctctatctga 780
caccagatgc gacctctat tccacttctc tgttcatctg ctgcttcttt tgttttaacca 840
gataaatctc cctcggggaa aaccgtcaaa aaaaaggcaa ctaaatgcaa acacgctcta 900
tagacaaaat gtgtttggtc tctgtcacga tggtgagaga gaattggcct ccgccgcaga 960
gatcgcttga ttattggcct ccagtgggca atgtcgggga aaaccaaact attgatgaga 1020
ggatcgacg aaaaatcaac aatgaccaac tttttgttac agttttgtta taaatatgag 1080
ttttggatat tccattgcgt atttttcttt tctactttca aaaaatctgc tccaaccttt 1140
aatggctttt cctgtcttgt caaaatctgg atttttgaat atataatttt taaaaccatc 1200
aaattcagcg aaatgaaatc atgtaataca attttttatt tttccgact gttgtgtatt 1260

```

ccatcaaaact	attcaaaaaa	tcaatataat	gatttttttt	tcattttttcg	gactctagag	1320
gatccccggg	gattggccaa	aggacccaaa	gggatgtttc	gaatgatact	aacataacat	1380
agaacatttt	caggaggacc	cttgcttggc	tagcaaaaaa	gcataagggt	ttgctggcac	1440
tgtcttttat	cttctctggc	ccagcatccg	cactggcagt	ctccgaaccg	gcctgcagga	1500
tcgatttttt	gcaaaattac	agcgtttag	ggggcggacg	ataggtccca	taggttttgc	1560
atatcatcat	tcattcattc	attggtacat	tcatttaccc	accttccctc	ttctgagctt	1620
ctctggagtt	ctgtgcttcc	tttttccctt	atctttatac	tgtaatTTTT	aactttcagg	1680
cattgattgg	atccccggga	ttggccaaag	gacccaaaag	tatgtttcga	atgatactaa	1740
cataacatag	aacattttca	ggaggacctt	tgcttgagg	gtaccgagct	cagaaaaaat	1800
gactgctcca	aagaagaagc	gtaaggtagc	ggtagaaaaa	atgagtaaag	gagaagaact	1860
tttcaactgga	gttgctccaa	ttcttggtga	attagatggg	gatgttaatg	ggcacaaatt	1920
ttctgtcagt	ggagagggtg	aagggtgatg	aacatacggg	aaacttaccc	ttaaatttat	1980
ttgcactact	ggaaaactac	ctgttccatg	ggtaagttta	aacatatata	tactaactaa	2040
ccctgattat	ttaaattttc	agccaacact	tgctcactact	ttctgttatg	gtgttcaatg	2100
cttctcgaga	taccagatc	atatgaaacg	gcctgacttt	ttcaagagtg	ccatgcccca	2160
aggttatgta	caggaaagaa	ctatatTTTT	caaagatgac	gggaactaca	agacacgtaa	2220
gtttaaacag	ttcgggtacta	actaaccata	catattttaa	ttttcagggtg	ctgaagtcaa	2280
gtttgaaggc	gatacccttg	ttaatagaat	cgagttaaaa	gggtattgatt	ttaaagaaga	2340
tggaaacatt	cttgagacaca	aattggaata	caactataac	tcacacaatg	tatacatcat	2400
ggcagacaaa	caaaaagaatg	gaatcaaatg	tgtaaagtta	aacatgattt	tactaactaa	2460
ctaattctgt	ttaaattttc	agaacttcaa	aattagacac	aacattgaag	atggaagcgt	2520
tcaactagca	gaccattatc	aacaaaatac	tccaattggc	gatggccctg	tccttttacc	2580
agacaaccat	tacctgtcca	cacaatctgc	cctttcgaaa	gatcccaacg	aaaagagaga	2640
ccacatggtc	cttcttgagt	ttgtaacagc	tgctgggatt	acacatggca	tggatgaact	2700
atacaaatag	cattcgtaga	attccaactg	agcgcgggtc	gctaccatta	ccaacttgtc	2760
tgtgtgcaaa	aataataggg	gccgctgtca	tcagagtaag	tttaaaactga	gttctactaa	2820
ctaacgagta	atatttaaat	tttcagcatc	tcgcgccgtg	gcctctgact	tctaagtcca	2880
attactcttc	aacatcccta	catgctcttt	cttccctgtg	tcccaccccc	tatttttgtt	2940
attatcaaaa	aaacttcttc	ttaatTTTct	tgTTTTtttag	cttcttttaa	gtcacctcta	3000
acaatgaaat	tgtgtagatt	caaaaataga	attaattcgt	aataaaaaagt	cgaaaaaaat	3060
tgtgtctcct	ccccccatta	ataataattc	tatcccaaaa	tctacacaat	gttctgtgta	3120
cacttcttat	gtttttttta	cttctgataa	atTTTTtttg	aaacatcata	gaaaaaacccg	3180
cacacaaaat	accttatcat	atgttacggt	tcagtttatg	accgcaattt	ttatttcttc	3240
gcacgtctgg	gcctctcatg	acgtcaaata	atgctcatcg	tgaaaaagtt	ttggagtatt	3300
tttggaattt	ttcaatcaag	tgaaagtTTa	tgaaattaat	tttctgctt	ttgctttttg	3360
gggggtttccc	ctattgtttg	tcaagagtTT	cgaggacggc	gtttttcttg	ctaaaatcac	3420
aagtattgat	gagcacgatg	caagaaagat	cggaagaagg	tttgggtttg	aggctcagtg	3480
gaagggtgag	agaagttgat	aatttgaaaag	tggagtagtg	tctatggggg	ttttgcctta	3540
aatgacagaa	tacattccca	atataccaaa	cataactggt	tcctactagt	cggccgtacg	3600
ggcccttttcg	tctcgcgctg	ttcggtgatg	acggtgaaaa	cctctgacac	atgcagctcc	3660
cggagacggg	cacagcttgt	ctgtaagcgg	atgccgggag	cagacaagcc	cgtcagggcg	3720
cgtcagcggg	tgttgccggg	tgctggggct	ggcttaacta	tgccgcatca	gagcagattg	3780
tactgagagt	gcaccatatg	cggtgtgaaa	taccgcacag	atgcgtaagg	agaaaaatcc	3840
gcacagggcg	gccttaaggg	cctcgtgata	cgcctatTTt	tataggttaa	tgtcatgata	3900
ataatggttt	cttagacgtc	aggtggcact	tttcggggaa	atgtgcgcgg	aacccttatt	3960
tgtttatttt	tctaaatata	ttcaaataatg	tatccgctca	tgagacaata	accttgataa	4020
atgcttcaat	aattttgaaa	aagggaagag	atgagtattc	aacatttccg	tgtcgccctt	4080
attccctttt	ttgcggcatt	ttgccttctt	gtttttgtct	accagaaaac	gctgggtgaa	4140
gtaaaagatg	ctgaagatca	gttgggtgca	cgagtgggtt	acatcgaact	ggatctcaac	4200
agcggtaaga	tccttgagag	ttttcgcccc	gaagaacggt	ttccaatgat	gagcactttt	4260
aaagtctctg	tatgtggcgc	ggtattatcc	cgtattgacg	ccgggcaaga	gcaactcggg	4320
cgccgcatac	actattctca	gaatgacttg	gttgagtact	caccagtcac	agaaaaagcat	4380
cttacgggat	gcatgacagt	aagagaatta	tgagtgctg	ccataacat	gagtataaac	4440
actgcggcca	acttacttct	gacaacgatc	ggaggaccga	aggagctaac	cgcttttttg	4500
cacaacatgg	gggatcatgt	aactcgccct	gatcgttggg	aaccggagct	gaatgaagcc	4560
ataccaaacc	acgagcgtga	caccacgatg	cctgtagcaa	tggcaacaac	gttgcgcaaa	4620
ctattaactg	gcgaactact	tactctagct	tcccggaac	aattaataga	ctggatggag	4680
gcggataaaag	ttgcaggacc	acttctgcgc	tcggcccttc	cggctggctg	gtttatttgc	4740
gataaatctg	gagccggtga	gcgtgggtct	cgcggtatca	ttgcagcact	ggggccagat	4800
ggtaaagccct	cccgtatcgt	agttatctac	acgacgggga	gtcaggcaac	tatggatgaa	4860

```

cgaaatagac agatcgctga gataggtgcc tcactgatta agcattggta actgtcagac 4920
caagtttact catatatact ttagattgat ttaaaacttc atttttaatt taaaaggatc 4980
taggtgaaga tcccttttga taatctcatg accaaaatcc cttaacgtga gttttcgttc 5040
cactgagcgt cagaccccggt agaaaagatc aaaggatcct cttgagatcc tttttttctg 5100
cgcgtaatct gctgcttga aacaaaaaaa ccaccgctac cagcgggtgt ttgtttgccg 5160
gatcaagagc taccaactct ttttccgaag gtaactggct tcagcagagc gcagatacca 5220
aatactgtcc ttctagtgtg gccgtagtta ggccaccact tcaagaactc tgtagcaccg 5280
cctacatacc tcgctctgct aatcctgtta ccagtggctg ctgccagtgg cgataagtcg 5340
tgtcttaccg ggttggactc aagacgatag ttaccggata aggcgcagcg gtcgggctga 5400
acgggggggt cgtgcacaca gccagcttg gagcgaacga cctacaccga actgagatac 5460
ctacagcgtg agcattgaga aagcgccacg cttcccgaag ggagaaaagg ggacaggat 5520
ccggtaaagc gcagggtcgg aacaggagag cgacagagg agcttccagg gggaaacgcc 5580
tggtatcttt atagtcctgt cgggtttcgc cacctctgac ttgagcgtcg atttttgtga 5640
tgctcgtcag gggggcggag cctatggaaa aacgccagca acgcggcctt tttacggttc 5700
ctggcctttt gctggccttt tgcctacatg ttcttctctg cgttatcccc tgattctgtg 5760
gataaccgta ttaccgcctt tgagtgaact gataccgctc gccgcagccg aacgaccgag 5820
cgacgcgagt cagtgaagca ggaagcggaa gagcgcccaa tacgcaaacc gcctctcccc 5880
gcgcgttggc cgattcatta atgcagctgg cacgacaggt ttcccgaact gaaagcgggc 5940
agtgaagcga acgcaattaa tgtgagttag ctactcatt aggcacccca ggctttacac 6000
tttatgcttc cggctcgtat gttgtgtgga attgtgagcg gataacaatt tcacacagga 6060
aacagctatg accatgatta cgccaagctg taagttaaaa catgatctta ctaactaact 6120
attctcattt aaattttcag agcttaaaaa tggctgaaat cactcacaac gatggatacg 6180
ctaacaactt ggaaatgaaa ta 6202

```

<210> 9

<211> 5737

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: plasmid
pGF2009

<400> 9

```

accctctatt ccacttctct gttcatctgc tgcctctttt gtttaaccag ataaatctcc 60
ctcgggggaaa accgtcaaaa aaaggcaaac taaatgcaaa cacgctctat agacaaaatg 120
tgtttgggtc cgtcacgaat ggtgagagag aattggcctc cgccgcagag atcgcttgat 180
tattggcctc cagtgggcaa tgcgggggaa aaccaaacta ttgatgagag gtatcgacga 240
aaaatcaaca atgaccaact ttttgttaca gttttgttat aaatatgagt tttggatatt 300
ccattgcgta tttttctttt ctactttcaa aaaatctgct ccaaccttta atggcttttc 360
ctgtcttgtc aaaatctgga tttttgaata tataattttt aaaaccatca aattcagcga 420
aatgaaatca tgaatacaaa ttttttattt ttcccgactg ttgtgtattc catcaacta 480
ttcaaaaaat caatataatg attttttttt carttttcgg actctagagg atccccgggg 540
attggccaaa ggacccaaag gtatgtttcg aatgatacta acataacata gaacattttc 600
aggaggaccc ttgcttggct agcaaaaatg cataagggtt tgctggcact gttctttatc 660
tttctggcac cagcatccgc actggcagtc tccgaaccgg cctgcaggat cgattttttg 720
caaattacga gcgttgtagg gggcggacga taggtcctat aggttttcta tatcatcatt 780
cattcattca ttggtacatt catttaccca ccttctctt tctgagcttc tctggagttc 840
tgtgcttctt ttttccctta tctttatact gtaattttta actttcaggc attgattgga 900
tccccgggat tggccaaagg acccaaaggt atgtttcgaa tgatactaac ataacataga 960
acattttcag gaggaccctt gcttggaggg taccgagctc agaaaaaatg actgctccaa 1020
agaagaagcg taaggtaccg gtagaaaaaa tgagttaaagg agaagaactt ttcactggag 1080
ttgtcccaat tcttgttgaa ttagatgggt atgtaaatgg gcacaaaatt tctgtcagtg 1140
gagagggtga aggtgatgca acatacggaa aacttaccct taaatttatt tctactactg 1200
gaaaaactacc tgttccatgg gtaagtttaa acatatatat actaactaac cctgattatt 1260
taaattttca gccacaactt gtcactactt tctgttatgg tgttcaatgc ttctcgagat 1320
acccagatca tatgaaacgg catgactttt tcaagagtgc catgcccga gggttatgtac 1380
aggaaagaac tatatttttc aaagatgacg ggaactacaa gacacgtaag tttaaacagt 1440
tcggtactaa ctaaccatac atattttaaat ttccaggtgc tgaagtcaag tttgaagggt 1500
atacccttgt taatagaatc gagttaaaaa gtattgattt taaagaagat ggaaacattc 1560

```

ttggacacaa	attggaatac	aactataact	cacacaatgt	atacatcatg	gcagacaaac	1620
aaaagaatgg	aatcaaagtt	gtaagttaa	acatgatttt	actaactaac	taatctgatt	1680
taaattttca	gaacttcaaa	attagacaca	acattgaaga	tggaagcgtt	caactagcag	1740
accattatca	acaaaatact	ccaattggcg	atgggctgtg	ccttttacca	gacaaccatt	1800
acctgtccac	acaatctgcc	ctttcgaaag	atcccaacga	aaagagagac	cacatgggtcc	1860
ttcttgagtt	tgtaacagct	gctgggatta	cacatggcat	ggatgaacta	tacaaatagc	1920
attcgtagaa	ttccaactga	gcgcgggtcg	ctaccattac	caacttgtct	ggtgtcaaaa	1980
ataatagggg	ccgctgtcat	cagagtaagt	ttaaactgag	ttctactaac	taacgagtaa	2040
tatttaaatt	ttcagcatct	cgcgcccggtg	cctctgactt	ctaagtccaa	ttactcttca	2100
acatccctac	atgctctttc	tccctgtgct	cccacccctt	atttttgtta	ttatcaaaaa	2160
aacttcttct	taatttcttt	gttttttagc	ttcttttaag	tcacctctaa	caatgaaatt	2220
gtgtagattc	aaaaatagaa	ttaattcgta	ataaaaagtc	gaaaaaaatt	gtgctccctc	2280
ccccatttaa	taataattct	atcccaaaat	ctacacaatg	ttctgtgtac	acttcttatg	2340
ttttttttac	ttctgataaa	ttttttttga	aacatcatag	aaaaaacccg	acacaaaata	2400
ccttatcata	tgttacgttt	cagtttatga	ccgcaatttt	tatttcttcg	cacgtctggg	2460
cctctcatga	cgtcaaatca	tgctcatcgt	gaaaaagttt	tggagtattt	ttggaatttt	2520
tcaatcaagt	gaaagtttat	gaaattaatt	ttcctgcttt	tgctttttgg	gggtttcccc	2580
tattgtttgt	caagagtttc	gaggacggcg	tttttcttgc	taaaatcaca	agtattgatg	2640
agcacgatgc	aagaaagatc	ggaagaaggt	ttgggtttga	ggctcagtg	aagggtgagta	2700
gaagttgata	atttgaaagt	ggagtagtgt	ctatgggggt	tttgccctaa	atgacagaat	2760
acattcccaa	tataccaaac	ataactgttt	cctactagtc	ggccgtacgg	gccctttcgt	2820
ctcgcgcgtt	tcggtgatga	cggtgaaaac	ctctgacaca	tgagctccc	ggagacgggtc	2880
acagcttgtc	tgtaagcgga	tgccggggagc	agacaagccc	gtcagggcgc	gtcagcgggt	2940
gttgccgggt	gtcggggctg	gcttaactat	gcggcatcag	agcagattgt	actgagagtg	3000
caccatattg	gggttgaaat	accgcacaga	tgctgaagga	gaaaataccg	catcaggcgg	3060
ccttaagggc	ctcgtgatac	gcctattttt	ataggttaat	gtcatgataa	taatgggttc	3120
ttagacgtca	ggtggcactt	ttcggggaaa	tgctgcggga	acccctattt	gtttattttt	3180
ctaaatacat	tcaaataatgt	atccgctcat	gagacaataa	ccctgataaa	tgcttcaata	3240
atattgaaaa	aggaagagta	tgagtattca	acatttccgt	gtcgccctta	ttcccttttt	3300
tgccggcattt	tgcccttctg	tttttgctca	cccagaaacg	ctggtgaaag	taaaagatgc	3360
tgaagatcag	ttgggtgcac	gagtgggtta	catcgaactg	gatctcaaca	gcggtaagat	3420
ccttgagagt	tttcgccccg	aagaacgttt	tccaatgatg	agcactttta	aagttctgct	3480
atgtggcgcg	gtattatccc	gtattgacgc	cgggcaagag	caactcgggtc	gccgcataca	3540
ctattctcag	aatgacttgg	ttgagtactc	accagtcaca	gaaaagcatc	ttacggatgg	3600
catgacagta	agagaattat	gcagtgctgc	cataaccatg	agtataaca	ctgcggccaa	3660
cttacttctg	acaacgatcg	gaggaccgaa	ggagctaacc	gcttttttgc	acaacatggg	3720
ggatcatgta	actcgccttg	atcgttggga	accggagctg	aatgaagcca	taccaaaccga	3780
cgagcgtgac	accacgatgc	ctgtagcaat	ggcaacaacg	ttgcgcaaac	tattaactgg	3840
cgaactactt	actctagctt	cccggaaca	attaatagac	tggatggagg	cggataaagt	3900
tgaggagcca	cctctgcgct	cggcccttcc	ggctggctgg	tttattgctg	ataaatctgg	3960
agccggtgag	cgtgggtctc	gcggtatcat	tgagcactg	gggccagatg	gtaagccctc	4020
ccgtatcgta	gttatctaca	cgacggggag	tcaggcaact	atggatgaac	gaaatagaca	4080
gatcgtgag	ataggtgcct	cactgattaa	gcattggtaa	ctgtcagacc	aagtttactc	4140
atatatactt	tagattgatt	taaaacttca	tttttaattt	aaaaggatct	aggtgaagat	4200
cctttttgat	aatctcatga	ccaaaatccc	ttaacgtgag	ttttcgttcc	actgagcgtc	4260
agaccccgta	gaaaagatca	aaggatcttc	ttgagatcct	ttttttctgc	gcgtaatctg	4320
ctgcttgcaa	acaaaaaaac	caccgctacc	agcggtaggt	tgtttgccgg	atcaagagct	4380
accaactctt	tttccgaagg	taactggctt	cagcagagcg	cagataccaa	atactgtcct	4440
tctagtgtag	ccgtagttag	gccaccactt	caagaactct	gtagcaccgc	ctacatacct	4500
cgctctgcta	atcctgttac	cagtggctgc	tgccagtggc	gataagtcgt	gtcttaccgg	4560
gttggaactca	agacgatagt	taccggataa	ggcgcagcgg	tcgggctgaa	cgggggggtc	4620
gtgcacacag	cccagcttgg	agcgaacgac	ctacaccgaa	ctgagatacc	tacagcgtga	4680
gcatttgagaa	agcggcacgc	ttcccgaagg	gagaaaggcg	gacaggtatc	cggtaagcgg	4740
cagggtcgga	acaggagagc	gcacgagggg	gcttccaggg	ggaaacgcct	ggtatcttta	4800
tagtctgtc	gggtttcgcc	acctctgact	tgagcgtcga	ttttgtgat	gctcgtcagg	4860
ggggcgaggc	ctatggaaaa	acgcccagcaa	cgcggccttt	ttacggttcc	tggccttttg	4920
ctggcctttt	gctcacatgt	tctttcctgc	gttatccctt	gattctgtgg	ataaccgtat	4980
taccgccttt	gagtgagctg	ataccgctcg	ccgcagccga	acgaccgagc	gcagcgagtc	5040
agtgagcgag	gaagcggaag	agcgcccaat	acgcaaaccc	cctctccccg	cgcgttggcc	5100
gattcatttaa	tgacgctggc	acgacaggtt	tcccagactg	aaagcgggca	gtgagcgcaa	5160

```

cgcaattaat gtaggttagc tcactcatta ggcaccccag gctttacact ttatgcttcc 5220
ggctcgtagt ttgtgtggaa ttgtgagcgg ataacaattt cacacaggaa acagctatga 5280
ccatgattac gccaagctgt aagtttaaac atgactctac taactaacta ttctcattta 5340
aattttcaga gcttaaaaat ggctgaaatc actcacaacg atggatacgc taacaacttg 5400
gaaatgaaat aagcttcatg cgtcttcttg attactgtag atgtttggca acttatcttt 5460
acataaaaaa gttttcattc attatttcca tcattcattt atctttctgt gtttttagtt 5520
agtttttagct agtttttttc taaattccta actttaaaaa atctggaaaa gaaaattaaa 5580
aaattttgtc cctattattt attttattac tggaaaatct tcaaacagga aaaccaccg 5640
cgttttgctt attgctgtat ttatgaaaaa aaaaacaatc aatattggtc aagtaataaa 5700
gaaaaaatta acgaatctct atctgacacc agatgcg 5737

```

<210> 10

<211> 6157

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: plasmid
pGF2013

<400> 10

```

ctcaccacaga aacgctgggtg aaagtaaaaag atgctgaaga tcagttgggt gcacgagtg 60
gttacatcga actggatctc aacagcggta agatccttga gagttttcgc cccgaagaac 120
gttttccaat gatgagcact tttaaagttc tgctatgtgg cgcggtatta tcccgatttg 180
acgcccgggca agagcaactc ggtcgccgca tacactatc tcagaatgac ttggttgagt 240
actcaccagt cacagaaaag catcttacgg atggcatgac agtaagagaa ttatgcagt 300
ctgccataac catgagtgat aacactgcgg ccaacttact tctgacaacg atcggaggac 360
cgaaggagct aaccgctttt ttgcacaaca tgggggatca tgtaactcgc cttgatcgtt 420
gggaaccgga gctgaatgaa gccataccaa acgacgagcg tgacaccacg atgctgtag 480
caatggcaac aacgttgcgc aaactattaa ctggcgaaact acttactcta gcttcccggc 540
aacaattaat agactggatg gaggcggata aagttgcagg accacttctg cgctcggccc 600
ttccggtctg ctggtttatt gctgataaat ctggagccgg tgagcgtggg tctcgcggt 660
tcattgcagc actggggcca gatggttaag cctcccgtat cgtagtatc tacacgacg 720
ggagtcaggc aactatggat gaacgaaata gacgatcgc tgagataggt gcctcactga 780
ttaagcattg gtaactgtca gaccaagttt actcatatat actttagatt gatttaaaac 840
ttcattttta atttaaaagg atctagggtga agatcctttt tgataatctc atgacaaaa 900
tcccttaacg tgagtttttcg ttccactgag cgtcagaccc cgtagaaaaa atcaaaggat 960
cttcttgaga tctttttttt ctgcgcgtaa tctgctgctt gcaaacaaaa aaaccaccg 1020
taccagcggg ggtttgtttg ccggtacaag agctaccaac tctttttccg aaggtaactg 1080
gcttcagcag agcgcagata ccaaatactg tctcttagt gtagccgtag ttaggccacc 1140
acttcaagaa ctctgtagca ccgcctacat acctcgtct gctaattcctg ttaccagtg 1200
ctgctgccag ttgcgataag tctgtcttta ccgggttga ctcaagacga tagttaccg 1260
ataaggcgca gcggtcgggc tgaacggggg gttcgtgcac acagcccagc ttggagcgaa 1320
cgacctacac cgaactgaga tacctacagc gtgagcattg agaaagcgcc acgcttccc 1380
aagggaagaa ggcgacagc tatccggtaa gcggcagggt cggaacagga gagcgcacga 1440
gggagcttcc agggggaaac gcctgggtatc tttatagtcc tgcgggttt cgccacctct 1500
gacttgagcg tcgatttttg tgatgctcgt cagggggggc gagcctatgg aaaaacgcca 1560
gcaacgcggc ctttttacgg ttccctggct tttgctggcc ttttgctcac atgttctttc 1620
ctgcttatc cctgattct gtggataacc gtattaccgc ctttgagtga gctgatacc 1680
ctcgcgcag ccgaacgacc gagcgcagcg agtcagtga cgaggaagcg gaagagcgcc 1740
caatacgcaa accgcctct cccgcgcgtt gcgcgattca ttaatgcagc tggcacgaca 1800
ggtttcccca ctggaaaagc ggcagtgagc gcaacgcaat taatgtgagt tagctcactc 1860
attaggcacc ccaggcttta cactttatgc ttcggctcg tatgttgtgt ggaattgtga 1920
gcgataaaca atttcacaca ggaaacagct atgaccatga ttacgccaag ctgtaagt 1980
aaacatgatc ttaactaacta actattctca tttaaatttt cagagcttaa aaatggctga 2040
aatcactcac aacgatggat acgctaacaa cttggaaatg aaataagctt gcatgcaagg 2100
ttaagtgccg gtttaagcat tatccccgca tgtagcttgt tcggcaccgt taaaaatgct 2160
gagtaatcag ctttttagaa tttaaaatat taaactttta aaattgcaac aaacatcgac 2220
aaatattcaa gaggcgaatg atatcgggaa ttctgattga aacgaaactg ttttgaaatt 2280
caaaaagtat tttcaaaagta ttgtccgcaa ggcacatcac gcaaacctgc agaactctacc 2340

```

gtatcccata	catttttata	gtttttccct	cagtttttaa	aattaaaaac	gctgaaaaag	2400
cgattaaatt	tattttaaatg	catcgttcga	ataaaataaa	gtttatcttt	tgataaaaaac	2460
atgagtttcc	tttggagaaa	agtagggatt	tgccttttca	aaaaattatt	tcgtgcagga	2520
tgctattttc	gtggcgaaac	ccatactcaa	gagctcatgc	gtcttcttga	ttactgtaga	2580
tgtttggcaa	cttattttta	cataaaaaacg	tttttattca	ttatttccat	cattcartta	2640
tctttctgtg	tttttagtta	gttttagcta	gtttttctct	aaattccctaa	ctttaaaaaa	2700
tctggaaaag	aaaattaaaa	aattttgtcc	ctattattta	ttttattact	ggaaaaatctt	2760
caaacaggaa	aaccacccgc	gttttgctta	ttgctgtatt	tatgaaaaaa	aaaacaatca	2820
atattgggtca	agtaataaag	aaaaaattaa	cgaatctcta	tctgacacca	gatgcgaccc	2880
tctattccac	ttctctgttc	atctgctgct	tcttttgttt	aaccagataa	atctccctcg	2940
gggaaaaccg	tcaaaaaaag	gcaaaactaaa	tgcaaacacg	ctctatagac	aaaatgtgtt	3000
tgggtctcgtc	acgaatggtg	agagagaatt	ggcctccgcc	gcagagatcg	cttgattatt	3060
ggcctccagt	gggcaatgtc	ggggaaaacc	aaactattga	tgagagggtat	cgacgaaaaa	3120
tcaacaatga	ccaacttttt	gttacagttt	tgttataaat	atgagttttg	gatattccat	3180
tgcgtatttt	tcttttctac	tttcaaaaaa	tctgctccaa	cctttaatgg	cttttctctg	3240
cttgtcaaaa	tctggatttt	tgaatatata	atttttaaaa	ccatcaaatt	cagcgaaatg	3300
aaatcatgta	atacaatttt	ttattttttc	cgactgttgt	gtattccatc	aaactcttca	3360
aaaaatcaat	ataatgattt	ttttttcatt	tttcggactc	tagaggatcc	ccggggattg	3420
gccaaaggac	ccaaagggtat	gtttcgaatg	atactaacat	aacatagaac	attttcagga	3480
ggacccttgc	ttggctagca	aaaatgcata	aggttttgct	ggcactgttc	tttatctttc	3540
tggcaccagc	atccgcactg	gcagtctccg	aaccggcctg	caggatcgat	tttttgcaaa	3600
ttacgagcgt	tgtagggggc	ggacgatagg	tccataggtt	tttgtatata	atcattcatt	3660
cattcattgg	tacattcatt	taccacett	cctcttctcg	agcttctctg	gagttctgtg	3720
cttccctttt	cccttatctt	tatactgtaa	tttttaactt	tcaggcattg	attggatccc	3780
cgggattggc	caaaggaccc	aaagggtatgt	ttcgaatgat	actaacataa	catagaacat	3840
tttcaggagg	acccttgcct	ggaggggtacc	ggtagaaaaa	atgagtaaa	gagaagaact	3900
tttctactgga	gttgtcccaa	ttcttgttga	attagatggt	gatgttaatg	ggcacaaatt	3960
ttctgtcagt	ggagaggggtg	aagggtgatgc	aacatacggg	aaacttacc	ttaaaattat	4020
ttgcactact	ggaaaaactac	ctgttccatg	ggtaagttta	aacatatata	tactaactaa	4080
ccctgattat	ttaaattttc	agccaacact	tgtcactact	ttctgttatg	gtgttcaatg	4140
cttctcgaga	taccagatc	atatgaaacg	gcatgacttt	ttcaagagtg	ccatgcccga	4200
aggttatgta	caggaaaagaa	ctatatTTTT	caaagatgac	gggaactaca	agacacgtaa	4260
gtttaaacag	ttcggactta	actaaccata	catattttaa	ttttcagggtg	ctgaagtcaa	4320
gtttgaaggt	gatacccttg	ttaatagaat	cgagttaaaa	ggatttgatt	ttaaagaaga	4380
tggaaacatt	cttgacaca	aattggaata	caactataac	tcacacaaatg	tatacatcat	4440
ggcagacaaa	caaaagaaatg	gaatcaaagt	tgtaaagtta	aacatgattt	tactaactaa	4500
ctaactctgat	ttaaattttc	agaacttcaa	aattagacac	aacattgaag	atggaagcgt	4560
tcaactagca	gaccattatc	aacaaaatac	tccaattggc	gatggccctg	tccttttacc	4620
agacaaccat	tacctgtcca	cacaatctgc	cccttcgaaa	gatcccaacg	aaaagagaga	4680
ccacatggtc	cttcttgagt	tgttaacagc	tgtcgggatt	acacatggca	tggatgaact	4740
atacaaatag	catctctaga	attccaactg	agcgccgggtc	gctaccatta	ccaacttgtc	4800
tgggtgtcaaa	aataataggg	gccgctgtca	tcagagtaag	tttaaaactga	gttctactaa	4860
ctaacgagta	atattttaa	tttcagcatc	tcgcgcccgt	gcctctgact	tctaagtcca	4920
attactcttc	aacatcccta	catgctcttt	ctccctgtgc	tcccaccccc	tatttttgtt	4980
attatcaaaa	aaacttcttc	ttaatTTCTT	tgttttttag	cttcttttaa	gtcacctcta	5040
acaatgaaat	tgtgtagatt	caaaaaataga	attaattcgt	aataaaaaagt	cgaaaaaaat	5100
tgtgctccct	gccccatta	ataataattc	tatcccaaaa	tctacacaat	gttctgtgta	5160
cacttcttat	gtttttttta	cttctgataa	attttttttg	aaacatcata	gaaaaaacccg	5220
cacacaaaat	accttatcat	atgttacgtt	tcagtttatg	accgcaattt	ttatttcttc	5280
gcacgtctgg	gcctctcatg	acgtcaaatc	atgctcatcg	tgaaaaagtt	ttggagtatt	5340
tttggaaatt	ttcaatcaag	tgaaagttta	tgaaattaat	tttctctgct	ttgctttttg	5400
gggttttccc	ctattgtttg	tcaagagttt	cgaggacggc	gtttttcttg	ctaaaatcac	5460
aagtattgat	gagcacgatg	caagaaagat	cggaagaagg	tttgggtttg	aggctcagtg	5520
gaaggtgatg	agaagtgtat	aatttgaaa	tggagttagt	tctatggggg	ttttgcctta	5580
aatgacagaa	tacattccca	atataccaaa	cataactgtt	tcctactagt	cgcccgtagc	5640
ggccctttcg	tctcgcgcgt	ttcgggtgat	acggtgaaaa	cctctgacac	atgcagctcc	5700
cggagacggg	cacagcttgt	ctgtaagcgg	atgccggggg	cagacaagcc	cgtcagggcg	5760
cgtcagcggg	tgttggcggg	tgtcgggggt	ggctttaaata	tgcggcatca	gagcagattg	5820
tactgagagg	gcaccatatg	cgggtgtgaaa	taccgcacag	atgcgtaagg	agaaaaatacc	5880
gcatacaggc	gccttaaggg	cctcgtgata	cgcctatttt	tatagggttaa	tgtcatgata	5940

```

ataatggttt cttagacgtc aggtggcact tttcggggaa atgtgcgcgg aaccttatt 6000
tggttttttt tctaaataca ttcaaatatg tatccgctca tgagacaata acctgataa 6060
atgcttcaat aatattgaaa aaggaagagt atgagatcc aacatttccg tgtcgcctt 6120
attccctttt ttgcggcatt ttgccttctt gttttttg 6157

```

<210> 11

<211> 5692

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: plasmid
pGF2014

<400> 11

```

taaacatata tatactaact aaccttgatt atttaaattt tcagccaaca cttgtcacta 60
ctttctgtta tgggtgttcaa tgcttctcga gatacccaga tcatatgaaa cggcatgact 120
ttttcaagag tgccatgccc gaaggttatg tacaggaaaag aactatattt ttcaaagatg 180
acgggaacta caagacacgt aagtttaaac agttcggtag taactaacca tacatattta 240
aattttcagg tgctgaagtc aagtttgaag gtgataccct tgtaataga atcgagttaa 300
aaggtattga ttttaaagaa gatggaaaaca ttcttggaca caaattggaa tacaactata 360
actcacacaa tgtatacatc atggcagaca aacaaaagaa tggaatcaaa gttgtaagtt 420
taaacatgat ttactaact aactaatctg atttaaattt tcagaacttc aaaattagac 480
acaacattga agatggaagc gttcaactag cagaccatta tcaacaaaat actccaattg 540
gcgatggccc tgtcctttta ccagacaacc attacctgtc cacacaatct gccctttcga 600
aagatcccaa cgaaaagaga gaccacatgg tccttcttga gtttgtaaca gctgctggga 660
ttacacatgg catggatgaa ctatacaaat agcattcgtg gaattccaac tgagcgccgg 720
tcgctaccat taccaacttg tctggtgtca aaaataatag gggccgctgt catcagatga 780
agtttaaaact gagttctact aactaacgag taatatttaa attttcagca tctcgcgccc 840
gtgcctctga cttctaagtc caattactct tcaacatccc tacatgctct ttctccctgt 900
gctcccaccc cctatttttg ttattatcaa aaaaacttct tcttaatttc tttgtttttt 960
agcttctttt aagtcacctc taacaatgaa attgtgtaga ttcaaaaata gaattaattc 1020
gtaataaaaa gtcgaaaaaa attgtgtctc ctccccccat taataataat tctatcccaa 1080
aatctacaca atgttctgtg tacacttctt atgttttttt tacttctgat aaattttttt 1140
tgaaacatca tagaaaaaac cgcacacaaa ataccttctc atatgttacg tttcagttta 1200
tgaccgcaat ttttatttct tcgcacgtct gggcctctca tgacgtcaaa tcatgctcat 1260
cgtgaaaaag ttttgagta ttttggaat ttttcaatca agtgaaagt tatgaaatta 1320
attttcctgc ttttgctttt tgggggtttc ccctattgtt tgtcaagagt ttcgaggacg 1380
gcgtttttct tgctaaaatc acaagtattg atgagcacga tgcaagaaag atcggaaaga 1440
ggtttggtt tgaggctcag tgggaagtg gtagaagttg ataatttgaa agtgagtag 1500
tgtctatggg gtttttgct taaatgacag aataacattc caatatacca aacataactg 1560
tttctacta tgcggccgta cgggcccttt cgtctcgcgc gtttcggtga tgacggtgaa 1620
aacctctgac acatgcagct cccggagacg gtcacagctt gtctgtaagc ggatgccggg 1680
agcagacaag cccgtcaggg cgcgtcagcg ggtgttggcg ggtgtcgggg ctggcttaac 1740
tatgcggcat cagagcagat tgtactgaga gtgcaccata tgcggtgtga aataccgcac 1800
agatgcgtaa ggagaaaata ccgcacacag cggccttaag ggctcgtga tacgcctatt 1860
tttataggtt aatgtcatga taataatggt ttcttagacg tcaggtggca cttttcgggg 1920
aaatgtgcgc ggaaccccta tttgtttatt tttctaaata cattcaaata tgtatccgct 1980
catgagacaa taaccttgat aaatgcttca ataatttga aaaaggaaga gtatgagtat 2040
tcaacatttc cgtgtcgcct ttattccctt ttttgcgga ttttgccctc ctgtttttgc 2100
tcaccacagaa acgctggtga aagtaaaaga tgctgaagat cagtgggtg cacgagtggg 2160
ttacatcgaa ctggatctca acagcggtaa gatccttgag agttttcgcc ccgaagaacg 2220
ctttccaatg atgagcactt ttaaagtctt gctatgtggc gcggtattat cccgtattga 2280
cgccgggcaa gagcaactcg gtcgcccac acactattct cagaatgact tgggtgagta 2340
ctcaccagtc acagaaaagc atcttacgga tggcatgaca gtaagagaat tatgcagtgc 2400
tgccataaacc atgagtgata acactgcggc caacttactt ctgacaacga tcggaggacc 2460
gaaggagcta accgcttttt tgcacaacat gggggatcat gtaactcgcc ttgatcgttg 2520
ggaaccggag ctgaatgaag ccataccaaa cgacgagcgt gacaccacga tgcctgtagc 2580
aatggcaaca acgttgcgca aactattaac tggcgaaacta cttactctag cttcccgga 2640
acaattaata gactggatgg aggcggataa agttgcagga ccacttctgc gctcggccct 2700

```

tccggctggc	tggtttattg	ctgataaatc	tggagccggt	gagcgtgggt	ctcgcggtat	2760
cattgcagca	ctggggccag	atggtaagcc	ctcccgtatc	gtagttatct	acacgacggg	2820
gagtcaggca	actatggtg	aacgaaatag	acagatcgct	gagatagggtg	cctcactgat	2880
taagcattgg	taactgtcag	accaagttta	ctcatatata	ctttagattg	atttaaaact	2940
tcatttttaa	tttaaaaagg	tctagggtgaa	gatccctttt	gataatctca	rgaccaaaat	3000
cccttaacgt	gagttttcgt	tccactgagc	gtcagacccc	gtagaaaaga	tcaaaggatc	3060
ttcttgagat	cttttttttc	tgcgcgtaat	ctgctgcttg	caaacaaaaa	aaccaccgct	3120
accagcgggtg	gtttgtttgc	cggatcaaga	gctaccaact	ctttttccga	aggttaactgg	3180
cttcagcaga	gcgcagatac	caaatactgt	ccttctagtgt	tagccgtagt	taggccacca	3240
cttcaagaac	tctgtagcac	cgccctacata	cctcgctctg	ctaactcctgt	taccagtggc	3300
tgtgcccagt	ggcgataagt	cgtgtcttac	cgggttggac	tcaagacgat	agttaccgga	3360
taaggcgcag	cggtcgggct	gaacgggggg	ttcgtgcaca	cagcccagct	tggagcgaac	3420
gacctacacc	gaactgagat	acctacagcg	tgagcattga	gaaagcgcca	cgcttcccga	3480
agggagaaaag	gcggacaggt	atccggtaag	cggcagggtc	ggaacaggag	agcgcacgag	3540
ggagcttcca	gggggaaacg	cctgggtatct	ttatagtccct	gtcgggtttc	gccacctctg	3600
acttgagcgt	cgatttttgt	gatgctcgct	agggggggcg	agcctatgga	aaaacgccag	3660
caacgcggcc	tttttacggt	tcctggcctt	ttgctggcct	tttgctcaca	tgctctttcc	3720
tgcgttatcc	cctgattctg	tggataaccg	tattaccgcc	tttgagttag	ctgataccgc	3780
tcgccgcagc	cgaacgaccg	agcgcagcga	gtcagttagc	gaggaagcgg	aagagcgccc	3840
aatacgcaaa	ccgcctctcc	ccgcgcgttg	gccgattcat	taatgcagct	ggcacgcacg	3900
gtttcccgac	tggaaagcgg	gcagttagcg	caacgcaatt	aatgtgagtt	agctcactca	3960
ttaggcaccc	caggctttac	actttatgct	tcgggtcgt	atgttgtgtg	gaattgtgag	4020
cggataacaa	tttcacacag	gaaacagcta	tgaccatgat	tacgccaaagc	tgtaagttta	4080
aacatgatct	tactaactaa	ctattctcat	ttaaattttc	agagcttaaa	aatggctgaa	4140
atcactcaca	acgatggata	cgctaacaac	ttggaaatga	aataagcttc	atgcgtcttc	4200
ttgattactg	tagatgtttg	gcaacttatt	tttacataaa	aacgttttca	ttcattattt	4260
ccatcattca	tttatctttc	tgtgttttta	gttagtttta	gctagttttt	ttctaaattc	4320
ctaactttta	aaaatctgga	aaagaaaaat	aaaaaatttt	gtccctatta	tttattttat	4380
tactggaaaa	tcttcaaaaca	ggaaaaccca	ccgcgttttg	cttattgctg	tatttatgaa	4440
aaaaaaaaca	atcaatattg	gtcaagtaaa	taagaaaaaa	ttaacgaatc	tctatctgac	4500
accagatgcg	accctctatt	ccacttctct	gttcatctgc	tgcttctttt	gtttaaccag	4560
ataaatctcc	ctcggggaaa	accgtcaaaa	aaaggcaaac	taaatgcaaa	cacgctctat	4620
agacaaaatg	tgtttggtct	cgtcacgaat	ggtgagagag	aattggcctc	cgccgcagag	4680
atcgcttgat	tattggcctc	cagtgggcaa	tgtcggggaa	aaccaaaacta	ttgatgagag	4740
gtatcgacga	aaaatcaaca	atgaccaact	ttttgttaca	gttttggtat	aaatatgagt	4800
tttgatattt	ccattgcgta	tttttctttt	ctactttcaa	aaaatctgct	ccaaccttta	4860
atggcttttc	ctgtcttgct	aaaatctgga	tttttgaata	tataattttt	aaaaccatca	4920
aattcagcga	aatgaaatca	tgtaatacaa	ttttttattt	tttccgactg	ttgtgtattc	4980
catcaaaacta	ttcaaaaaat	caatataatg	attttttttt	catttttcgg	actctagagg	5040
atccccgggg	attggccaaa	ggacccaaag	gtatgtttcg	aatgatacta	acataacata	5100
gaacattttc	aggaggaccc	ttgcttggct	agcaaaaatg	cataagggtt	tgttggcact	5160
gttctttatc	tttctggcac	cagcatccgc	actggcagtc	tccgaaccgg	cctgcaggat	5220
cgattttttg	caaattacga	gcgttgtagg	gggcggaacg	taggtcctat	aggttttgta	5280
tatcatcatt	cattcattca	ttggtacatt	catttaccca	ccttcctctt	tctgagcttc	5340
tctggagttc	tgtgcttccct	ttttccctta	tctttatact	gtaattttta	acttttcaggc	5400
attgattgga	tccccgggat	tggccaaagg	acccaaagggt	atgtttcgaa	tgataactaac	5460
ataacataga	acattttcag	gaggaccctt	gcttggagggt	taccggtaga	aaaaatgagt	5520
aaaggagaaag	aacttttcac	tggagttgtc	ccaattcttg	ttgaattaga	tgggtgatgtt	5580
aatgggcaca	aattttctgt	cagtggagag	ggtgaagggt	atgcaacata	cggaaaaactt	5640
acccttaaat	ttatttgcac	tactggaaaa	ctacctgttc	catgggtaag	tt	5692

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
21 September 2000 (21.09.2000)

PCT

(10) International Publication Number
WO 00/54815 A3

(51) International Patent Classification⁷: **C12N 15/12**,
15/85, 5/10, C07K 14/435, A01K 67/033, C12Q 1/68,
A61K 49/00

Technologiepark 9, B-9052 Zwijnaarde (BE). ROELENS, Ingele [BE/BE]; Devgen NV, Technologiepark 9, B-9052 Zwijnaarde (BE). BOGAERT, Thierry [BE/BE]; Devgen NV, Technologiepark 9, B-9052 Zwijnaarde (BE).

(21) International Application Number: **PCT/EP00/02373**

(22) International Filing Date: **16 March 2000 (16.03.2000)**

(74) Agents: **BAVERSTOCK, Michael, George, Douglas et al.**; Boulton Wade Tennant, Verulam Gardens, 70 Gray's Inn Road, London WC1X 8BT (GB).

(25) Filing Language: **English**

(26) Publication Language: **English**

(30) Priority Data:
9906018.8 **16 March 1999 (16.03.1999) GB**

(81) Designated States (*national*): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(71) Applicant (*for all designated States except US*): **DEVGEN NV [BE/BE]**; Technologiepark 9, B-9052 Zwijnaarde (BE).

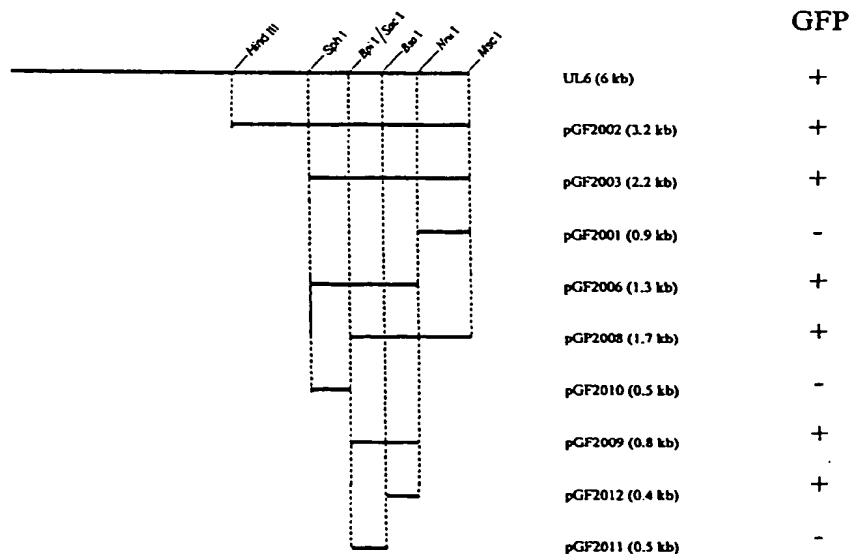
(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **ZWAAL, Richard [BE/BE]**; Devgen NV, Technologiepark 9, B-9052 Zwijnaarde (BE). **ASAERT, Wouter [BE/BE]**; Devgen NV,

[Continued on next page]

(54) Title: **EXPRESSION OF DNA OR PROTEINS IN *C. ELEGANS***



(57) Abstract: DNA fragments from the promoter region of the *C. elegans* UL6 gene which are capable of functioning as promoters directing gene expression in the excretory cell of *C. elegans* are provided and also expression vectors and transgenic *C. elegans* containing these fragments. Also provided are screening methods performed in *C. elegans* for identifying compounds or mutations which have an effect on the morphology of the excretory canal. Compounds identified using these screening methods may have therapeutic potential in the treatment of a range of diseases for which the *C. elegans* excretory canal serves as a model.

WO 00/54815 A3



Published:

— *With international search report.*

(88) Date of publication of the international search report:
25 January 2001

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 00/02373

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/12 C12N15/85 C12N5/10 C07K14/435 A01K67/033
C12Q1/68 A61K49/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C07K A01K C12Q A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	YOUNG J AND HOPE I: "Molecular markers of differentiation in <i>Caenorhabditis elegans</i> obtained by promoter trapping." DEV DYN., vol. 196, no. 2, February 1993 (1993-02), pages 124-132, XP002146202 cited in the application page 128, column 2, paragraph 6 -page 129, column 1, paragraph 4; table 2	1-14
X	BROEKS A ET AL: "A P-glycoprotein protects <i>Caenorhabditis elegans</i> against natural toxins" EMBO JOURNAL., vol. 14, no. 9, 1 May 1995 (1995-05-01), pages 1858-1866, XP002146203 figures 3-5	11,13, 25,26, 36, 41-43, 47,48
	--- -/--	

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

30 August 2000

Date of mailing of the international search report

15/09/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Lonnoy, O

INTERNATIONAL SEARCH REPORT

Int l Application No

PCT/EP 00/02373

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LAMBIE E ET AL: "Two homologous regulatory genes, lin-12 and glp-1, have overlapping functions" DEVELOPMENT, vol. 112, no. 1, May 1991 (1991-05), pages 231-240, XP000938910 page 235, column 2, paragraph 2 -page 238, column 1, paragraph 1	19,20, 23,55,56
A	C. ELEGANS SEQUENCING CONSORTIUM: "Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium" SCIENCE., vol. 282, no. 5396, 11 December 1998 (1998-12-11), pages 2012-2018, XP002146204 cited in the application	1-6
A	-& DATABASE EM-INV E.B.I., Hinxton, U.K.; Accession Number: AF045642, 11 February 1998 (1998-02-11) GEISEL C: "The sequence of C. elegans cosmid C17H12" XP002146205 Nucleotides 10576-13785: 100% identity in 3210 nt overlap with SeqIdNo.2 abstract	
A	NELSON F AND RIDDLE D: "Functional study of the Caenorhabditis elegans secretory-excretory system using laser microsurgery" J EXP ZOOL, vol. 231, no. 1, July 1984 (1984-07), pages 45-56, XP000938887	
A	WO 97 11956 A (UNIV COLUMBIA ;GREENWALD IVA (US); LEVITAN DIANE (US)) 3 April 1997 (1997-04-03) page 18, line 26-29	
A	WO 98 28971 A (LINK CHRISTOPHER ;UNIV TECHNOLOGY CORP (US)) 9 July 1998 (1998-07-09)	
A	WO 96 38555 A (BOGAERT THIERRY ;STRINGHAM EVE (CA); VANDEKERCKHOVE JOEL (BE)) 5 December 1996 (1996-12-05) page 57, line 17 - line 20	
A	WO 98 24810 A (BOGAERT THIERRY ANDRE OLIVIER ;DERAEYMAEKER MARC (BE); VANDEKERCKH) 11 June 1998 (1998-06-11)	

FURTHER INFORMATION CONTINUED FROM PCT/SA/ 210

Continuation of Box I.2

Claims Nos.: 30-35, 49-54

Claims 30-35 and 49-54 relate to compounds characterised by a desirable property, namely that they are modulators or otherwise inhibitors or enhancers of growth cone steering or one of the other listed physiological processes, and to applications of said compounds, without giving a true technical characterization. Moreover, no such compound is defined in the application. In consequence, the scope of said claims is ambiguous and vague, and their subject-matter is not sufficiently disclosed and supported (Art. 5 and 6 PCT). No search can be carried out for such speculative claims the wording of which is, in fact, a recitation of the results to be achieved.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/EP 00/02373

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
W0 9711956 A	03-04-1997	AU 7251496 A CA 2233297 A EP 0854881 A JP 11512611 T	17-04-1997 03-04-1997 29-07-1998 02-11-1999
W0 9828971 A	09-07-1998	AU 5619298 A EP 0948612 A	31-07-1998 13-10-1999
W0 9638555 A	05-12-1996	AU 6123496 A EP 0832222 A	18-12-1996 01-04-1998
W0 9824810 A	11-06-1998	AU 5662298 A EP 0941239 A	29-06-1998 15-09-1999